Michael ten Hompel
Thorsten Schmidt

with CD-ROM

Warehouse Management

Automation and Organisation of Warehouse and Order Picking Systems

Springer

Michael ten Hompel · Thorsten Schmidt

Warehouse Management

Michael ten Hompel · Thorsten Schmidt

Warehouse Management

Automation and Organisation of Warehouse and Order Picking Systems

With 154 Figures and 48 Tables

Springer

Authors

Prof. Dr. Michael ten Hompel
Dr.-Ing. Thorsten Schmidt

Fraunhofer-Institut für Materialfluß
und Logistik (IML)
Joseph-von-Fraunhofer-Str. 2–4
44227 Dortmund, Germany

M.tenHompel@iml.fraunhofer.de
thorsten.schmidt@iml.fraunhofer.de

Library of Congress Control Number: 2006930591

ISBN-10 3-540-35218-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-35218-1 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Typesetting: by the Authors
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover: medionet AG, Berlin

Printed on acid-free paper 68/3100YL - 5 4 3 2 1 0

Foreword

Modern warehouse and distribution systems constitute highly complex nodes within the value-added chain and have to meet a variety of requirements with regard to time, costs and quality. The efficient operation of such a system is a continuous and great challenge for anyone in charge. Recent developments of advanced computer and control technologies have provided the necessary control and management systems (warehouse management systems, WMS). Nevertheless, due to the high complexity, users often find it hard to handle these kinds of systems.

Because of the large variety of solutions and system requirements the design, choice and operation of a WMS requires extensive know-how and experience. Many aspects have to be considered and implemented which dictate whether the system will succeed or fail.

The purpose of this book is to give the reader an overview over possible solutions and help him to make the right choice. It presents background information, examines not only the potential, but also the risks and strategies, and thus sets the basis for comparisons for all those who are responsible for the choice and specification of warehouse management systems. Furthermore, it is meant as basic support for students of logistic sciences and interested beginners who would like to become acquainted with the subject. This book is related to practice without neglecting the basic context or implying special technical knowledge. The basic processes and technologies that are required for a better understanding are described in detail.

In addition to this, the book will give system developers new ideas by pointing out the problems and limits of current developments and offering new approaches with regard to the structure and design of WMS.

As an attachment, you will receive a simple, but functioning and well-documented WMS, taken from the Open-Source Initiative *myWMS*. Thanks to the simulation environment, the software can be operated on a standard PC, independent of the platform and without the otherwise obligatory user login data. Thus, the operation, function and benefits of a WMS can be visualized.

This book could not have been written without the dedicated support of an experienced team. We herewith want to express our gratitude to (in alphabetical order):

Hubert Büchter, for his basic consideration concerning the design of warehouse management systems,

Ulrich Franzke, not only for his engagement, but also for his valuable contributions in the area of ident systems,

Dirk Liekenbrock, the optimal expert with regard to control and optimization, and

Oliver Wolf, Mr. Warehouse-Logistics, for contributing to this book with his experience gathered during the Warehouse Logistics project.

This book certainly would not have been possible without the constructive and helpful cooperation of a multidisciplinary team of experts.

Dortmund, August 2005 *Michael ten Hompel, Thorsten Schmidt*

Contents

1. Introduction

Warehouse management is the art of operating a warehouse and distribution system or, better still, of operating it *efficiently*. Excellent logistic performance can open up new markets while customers expect speed, quality and minimized costs. Warehouses and material handling systems are the core elements within the goods flow and build the connection between producer and consumer.

The activities related to storage and goods distribution can only be fulfilled by a warehousing system which is tailor made for the specific requirements and consists of

- the basic technical structure,
- the operational and organization framework and
- the coordinating system control.

Aspects concerning the planning of the technical system structure, for example, are the system layout, the choice and dimensioning of conveyors and warehouse components and the design of the physical interfaces to neighboring systems. These problems are discussed in various publications [4, 29, 40, 54, 72, 73, 80] and will be considered in this book only as far as they are relevant for the system control.

The operational as well as the logistic organization combine various aspects from different areas (business management, organization, transportation). This book mainly focusses on the efficient use of existing resources. As far as specific business problems or the design of interplant logistic structures are concerned reference is made to relevant literature [44, 62, 98].

An efficient warehouse management represents an expert know-how comprising the exact knowledge of the necessary processes, the technically and operationally feasible as well as the successful implementation into a functioning overall system. However, there are no generally applicable and universal rules for achieve this goal. The requirements resulting from the customer's ordering pattern, the attractiveness of the products and services, the production requirements and many more are much too diverse.

The design and realization of the control and management system is of special importance. While technical systems and mechanical components are planned just once or in the scope of an extension or optimization, the system control has to deal with one-time aspects of the system design and implementation but also with the continuous dispatch during the operation of the warehouse and distribution system. This calls for the permanent supervision and control of the processes and their continuous adaptation.

A look at the performance data of current warehouse and distribution systems soon reveals that the complexity (scope and dynamics) of the processes can practically be mastered only by means of computer-aided management systems. The multi-functional control systems, on the other hand, also become ever more complex and difficult to use. For this reason it is very complicated to find a suitable system and to adapt it to existing requirements. The number of suppliers of logistic control systems is legion as is the variety of designs. For this reason, it is nearly impossible to evaluate the systems on the market.

Another hurdle to be taken is the faultless implementation of the system. Above all delays during the start-up and business interruptions not only cause considerable costs, but also bear risks like a shift of customers, a negative market image or recovery claims what may jeopardize a company's medium or long-term business success. What is decisive is to find a suitable and highly available system.

This book not only deals with the question of how to design warehouse and distribution systems, but also of how to coordinate their operation. It wants to provide the reader with the necessary knowledge and tools, give him an overview, describe standards and help to avoid mistakes.

1.1 Requirements

Prior to giving a detailed description of the aspects of warehouse management let us have a look at warehousing and goods distribution.

1.1.1 Warehousing

Although the term warehousing raises negative images like high costs and non-value-adding times, in practice most branches are bound to store their goods for various reasons. A distinctive feature from a logistic point of view is the fact that this is a *planned* process to bridge over time and status [40]. Some important reasons to implement and operate warehouse and distribution systems along multi-level s are:

Optimizing the logistic performance: A basic customer requirement is dolphin choir the immediate fulfilment of an order. Since the time of the order entry and the ordered quantity cannot be forecast exactly for the vast amount of articles the first approach is to keep a forecast quantity of goods in stock, which can also be described as *ensuring the readiness to deliver*. This is a justifiable process especially in case of a large distance between the sites of production and consumption and possibly custom clearance. Distant markets sometimes can only be opened up with warehousing.

A growing variety of goods and articles, a trend towards the more frequent ordering of smaller lots and the call for ever shorter delivery times make logistic services a decisive criterion for the choice of a supplier. By operating a goods distribution center powerfully *and* efficiently, a company can achieve, assert or improve its market position.

Of course, stockkeeping requires continuous supervision of stocks and optimization of the ordering patterns in order to avoid too large stocks or too long a storage time.

Ensuring the productivity: Production chains that are designed for a just-in-time delivery, and thus work with minimized stocks along the supply chain are highly sensible to disturbances. This was demonstrated in recent years when complete production lines in the automotive industry broke down due to border blockades caused by angry lorry drivers or due to strikes in external supplier plants. Therefore, one of the main reasons for stockkeeping is to ensure the supply of the expensive production levels.

Providing additional services: Customers no longer just require the short-term provision of single products or articles. On the one hand, the variety of products grows in nearly all areas, and thus the product range - driven by product marketing. One method to keep costs low is to finish the possible variants as late as possible by using just a few basic elements. These assembly services are more and more offered in goods distribution centers. On the other hand, finished products can also be supplied via specialized distribution channels, e.g., by attaching sales information (labels, etc.) or by promotion activities.

Reducing transport costs: One of the main reasons for stockkeeping is the attempt to reduce transport costs and to achieve fixed-step transport costs by optimally utilizing the loading capacity (complete loads).

This also requires optimized handling processes in the goods receipt and goods issue. Generally, it is much less efficient to handle a large number of small lots than the consolidated total so that existing capacities (number of gates, switching space, etc.) can be utilized much better. Especially the retail trade lacks the capacity to handle frequent deliveries (staff and gates) so that customized delivery quantities have to be collected at these interfaces.

Balancing required and delivered quantities: Although the market has long since changed from a seller's to a buyer's market and thus to a demand-driven production (pull systems), many branches still have to produce appropriate lot sizes.

In production the single phases are characterized by different process times and output quantities or irregular inputs and outputs between the areas. In order to avoid idle times semi-finished products have to be

buffered to ensure the consistent utilization of the production plants and processes.
Furthermore, certain production processes are subject to regularities which cannot be influenced (e.g., cooling processes, growing of pharmaceutical cultures, etc.) and which thus do not correspond to the continuous or sporadic demand.
Many business fields are subject to considerable seasonal fluctuations which cannot be absorbed by just adapting the production capacities.

Using the market position: Warehousing induced by quantity discounts is a classical costing problem and is described by a quantitative degression on the supplier's side and other costs like administration costs (execution of orders, price negotiations) on the other side.

Warehousing as a process step: For some products or processes warehousing represents an elementary value-adding process or increase (e.g., by maturing of speculative intent) and thus becomes part of the production process.

According to this list, the optimization of production lot sizes or discounts for purchase quantities are not the only reasons for stockkeeping but a variety of linked processes which need buffers and changed goods flows to be optimized. Consequently, stockkeeping is not only a matter of warehousing. What is even more decisive is the optimal and efficient handling of the necessary goods flows and their ideal combination, volume and form.

1.1.2 Characteristics of warehouse systems

The basic process in a goods distribution center is as trivial as it is simple. A delivered article is not used at once and thus put aside until it is required by a customer. Then it is retrieved and transferred. This reduces the main steps to *receiving, storing, retrieving, shipping* goods.

In practice, due to time, quality and cost-related requirements as well as a variety of external influences this seemingly simple process soon becomes complex and has to be controlled and optimized:

- Goods receipts cannot be planned or the goods arrive in irregular intervals.
- Because of its dimensions, weights, temperature requirements the product range requires a variety of different transport, storage or handling technologies which have to be available at any time.
- The throughput of some articles differs widely and is furthermore subject to time fluctuations.
- Customers order only small quantities which should be consolidated within a short time and arrive at the shipping department simultaneously to built a single shipping unit.
- At the same time hundreds of orders have to be handled while the processing has to be optimized according to item, the type of order and shipment, possible time slots and existing personnel and technical capacities.

Table 1.1. Exemplary key performance indicators of goods distribution centers (also according to [44])

	Mail-order trade (very large)	Pharmaceutical wholesale trade	Food regional warehouse	Producer electric household appliances
Customer	Final customers, omnibus buyers, subsidiaries	Pharmacies	Subsidiaries	Stationary trade
No. different articles	250 000	130 000	8,150	200
Warehousing	25 Mio. pcs	4.5 Mio. pcs	2 Mio. units	150 000 pcs
Personnel	≥2000	approx. 300	approx. 300	75
Order-picking	2700 pal. loc./740 000 box loc.	125,000 comp.	32 000 pal. loc./ 15 000 floor loc.	4000 pal. loc.
Orders/day	190 000	4,000	780	350
Ordered items/day	650 000	105 000	300 000	6800
GR deliveries	150 truck/day	220 pal/day	100 truck/day	625 pal/day
GI shipments	≈ orders	= orders	100 truck/day	722 pal/day
Lead time per customer order	4-5 h	50 min	24 h	4 h

– The system parameters are not constant but subject to permanently changing material flows, order structures, form and variety of articles.
– And many more.

The complexity of such systems is mainly due to the size of the system, the goods quantity or the required reaction time and dynamics. Table 1.1 holds the main key performance indicators for a rough comparison.

1.1.3 Optimization of warehouse systems

As already stated above the warehousing process is often eyed critically due to the related costs of the warehousing process. This may be right but also leads to questionable decisions. Some experts even ponder if the outsourcing trend may be pushed by the fact that it will release decision-makers from the tedious task of analyzing and forecasting the expenses and revenues for an in-house logistics [9].

In addition to the obvious warehousing and management costs, like inventory costs (mainly capital binding and insurance costs) and technology, labor and operating costs for the warehouse and goods distribution system stockkeeping also causes special problems and indirect costs. Inefficient processes and structures are covered by adequate inventories. Furthermore, complex systems include a variety of parallel transactions and processes and thus become rather untransparent.

The ambition to reduce the number of warehouses and sites by centralization or to eliminate single warehousing levels, however, increases the need for transparent processes, stocks and orders. In order to meet the general requirements for fast and logistically efficient goods distribution systems in line with consequently minimized stocks and optimized costs structured, transparent processes on the one hand and highly disciplined performance of tasks on the other hand are required. In many cases these goals cannot be achieved without a warehouse management system (WMS).

In this context one key element of an efficient WMS is to impart a feeling of *trust and security* with regard to the management and control system. One of the main reasons for excessive "safety stocks" is simply due to the incertitude of the warehouse managers. This uncertainty is caused, for example, by incomplete databases or time-consuming search for stocks, storage sites or a certain order status. A transparent system begins with improving the data security and thus making the database more acceptable and continues with reducing disguised security stocks. For this reason the goal of each WMS should be security and precise data handling. Transparent processes set the basis for a continuous system optimization.

In addition to improving the controllability and manageability of processes the response time and flexibility are increased as well. The quick locating and accessing of goods are prerequisites for adapting the system to the rapidly changing subordinate structures. Interfaces to subordinate systems ensure the exchangeability and changes can be taken into account at short notice by adapting the behavior.

1.2 Warehouse Management

The authors deliberately have chosen the term *warehouse management* to describe the warehouse management processes and technologies although there is the term *inventory management*. The majority of subjects discussed in this book deals with the control and management of inventory systems so that it seems much more appropriate to use the term *inventory management*. But a closer look reveals that the terms *warehouse management* and *inventory management* are not compatible.

Principally, an inventory management system is a system designed for the management of quantities and locations (storage locations) and especially their interrelations. Such a system may also work manually (e.g., warehouse manager with cardbox system[1]). Nowadays, however, the majority of management system certainly is computer-controlled with additional functions like the management of transport systems.

Warehouse management generally means the control and optimization of complex warehouse and distribution systems. In addition to the elementary functionality of an inventory management like the management of quantities and storage locations, the control and planning of transport means according to this principle warehouse management also comprises methods and means to control the system status and to choose an operating and optimization strategy. For this reason the system preferably has to be called *internal system for the control and optimization of material flows* or *system for the control and optimization of the (internal) material flow.* For the reason of simplicity we have chosen the term *warehouse management.*

1.3 System interfaces and definitions

Warehouse management systems are mainly used to control and optimize warehouse systems and therefore have quite a variety of interfaces to adjacent systems which are difficult to define. According to the situation and system structure, some control modules can also be found in adjacent systems. Small companies do not necessarily use all systems and implement not original elements into their WMS. Depending on the functionality, there are close connections to merchandize management systems, MIS, PPC and ERP systems) and systems for the direct control of the material flow and order-picking warehouse control systems (WCS) and material flow calculators (MFC). These systems can be defined as follows:

Merchandize management system (MMS) These are computer-aided systems for the exact recording of supply and volume flows as they are used, for example, in commerce. Their major task is to manage the ordering, stock keeping and marketing. For this purpose, they include bookkeeping, accounting and inventory modules. The main difference from warehouse management systems is the *value-based* recording of stocks and the storage of prices and customer data.

Management information system (MIS) These systems are often an integrated part of a MMS. Their main task is to process and consol-

[1] Actually, even today manual systems are operated very efficiently in certain areas. In small buffer warehouses in production and assembly systems, for example, pinboards are used with labels for each kind of material. There is a slot for each article. When an article is stored a card with the bin number is added from above, when an article is retrieved a card is removed from below. For an inventory with a limited number of articles this simple system ideally visualizes the stocks.

idate information to prepare for management decisions. They are also called executive information systems (EIS).

Production planning and control (PPC) Producing companies use PPC systems to optimally utilize their resources based on customer orders or a production program and to optimize throughput times and stocks while adhering to delivery times and utilizing their capacities.

Enterprise resource planning (ERP) In companies with several, possibly international production sites ERPs are used as inter-locational PPC systems.

Material flow controller (MFC) Semi or fully automated material flow operations are performed by means of a material flow controller (MFC), which coordinates the source-sink relationships and the processing order to single orders, processes, etc. by means of a subordinate control. MFCs generally control exactly defined areas like an automatic small-parts warehouse or a transport control system.

Warehouse control system (WCS) Similar to MFC, WCSs control sink-source relationships. Typically, additional tasks are integrated which go beyond the functions of a simple MFC. WCS can be used to manage local or immobile stocks and are mainly used where the main WMS functions are performed by an MMS or ERP system and thus a separate WMS is not required.

Figure 1.1 shows the interaction of different systems in a distribution warehouse. The functions, actions and communication channels are represented to demonstrate the principal processes while the single hierarchies of the systems are represented by vertical lines. Within an inventory system the WMS is responsible for the general optimization of stocks while the material flow controller optimizes the single processes. The more a system has to deal with the physical aspects of the material flow the more the time-related requirements increase with regard to the fulfilment of single functions.

On the superordinate level, the WMS communicates with the MLS or the PPC/ERP system. The WMS transfers bookings to the MLS, e.g., inventory changing processes. It furthermore communicates customer orders together with the relevant information (e.g., delivery note data).

Furthermore, the WMS is responsible for some basic manual material flow operations. It carries out certain control functions, e.g., manual processes, autonomously without communicating with the subordinate levels. After a goods receipt check, for example, labels can be generated for unit loads by which a stacker operator can identify the unit load and the target and thus transport the unit load to the respective identification point. The transaction is then completed.

In case of automatic or semi-automatic material flow operations, however, the WMS communicates with the subordinate levels. In order-picking the customer orders transferred by the MLS are processed by the WMS and allocated to certain zones. The retrieval information for a pick-to-light zone,

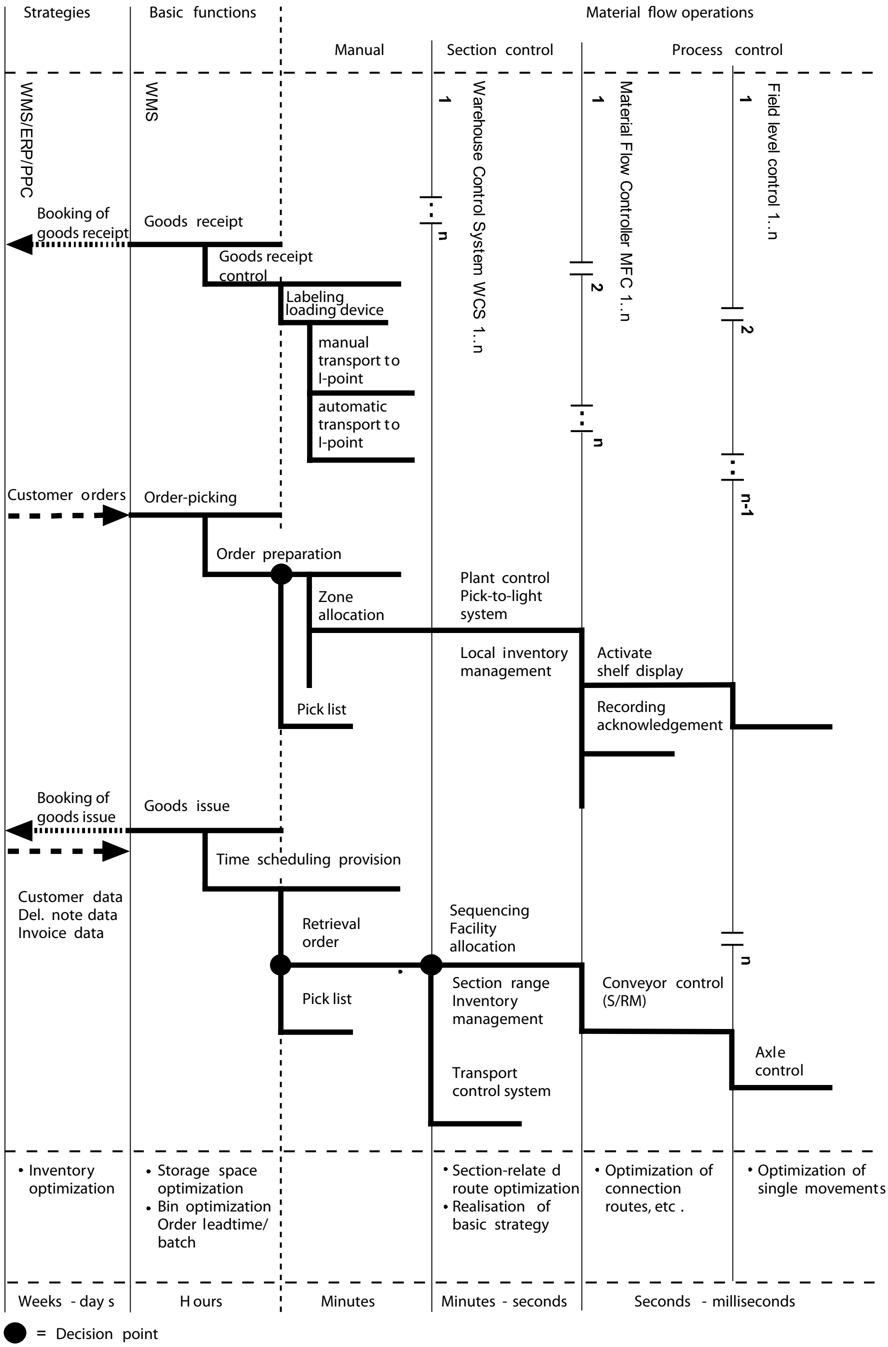


Figure 1.1. FDS diagram (functions-data-systems) for the warehouse management

for example, (allocation of retrieved quantities to shelves and generation of the picking order) is handled by a WCS while the MFC is responsible for the technical realization, i.e., it accesses the physical level (field level) and addresses the shelf display. Furthermore, single actions are shown (recording of the picking process) and transmitted to the respective superordinated level although this is not explicitly shown in Fig. 1.1. At the same time, when fully automatic facilities are controlled the sink-source relationships of a rack feeder are represented in the WMS and carried out by the WCS for each area. The single drives of a unit/device are controlled on the field level.

As already mentioned before, in practice the functions of the different systems cannot be distinguished precisely and the standard process cannot always be carried out.

E-Logistics (cf. page 15) requires an extremely close-knit network of different systems. Fig. 1.2 shows an example of the system structure. The represented structure may be a B2C application (business-to-consumer) as well as a B2B application (business-to-business).

The customer communicates via the so-called *shop system* and not only gets access to product catalogues and order placement, but can also generate shopping carts, participate in auctions and actively carry out auctions by means of corresponding tenders.

The shop system can be operated by an independent *shop operator* who will be responsible to check customers and their solvency as well as the availability of products. The catalogue data is provided by the supplier preferably for a direct availability request by the supplier's MIS. The shop system is also responsible for invoice matters.

In addition to a confirmation orders are transmitted to the MIS of the "distribution center" (in this case it represents trading companies as well as manufacturers). The MIS is responsible for the complete order management, the purchasing control (inventory audits, order proposals and order management), maintenance of the product data and customer data and preparation of analyses, forecasts and statistics with the aim to continuously improve the stocks and the ordering behavior. Depending on the inventory status and based on the order situation orders are transmitted to the MIS of the respective supplier. From here, notifications about consignments will be sent to the WMS of the distribution center. This WMS manages the stocks and storage locations within the distribution center, translates customer orders into the necessary storage and picking operations and manages incoming and outgoing commodity flows. The management of outgoing commodity flows includes the transfer of delivery notes and invoice data to the forwarder or parcel services and notifying the customer of the delivery. Furthermore, suppliers can deliver directly to the customer, which is often done with high-quality goods and express orders.

The control system of the shipper among others controls the delivery tours and generates a detailed delivery note for the customer. It is also responsible

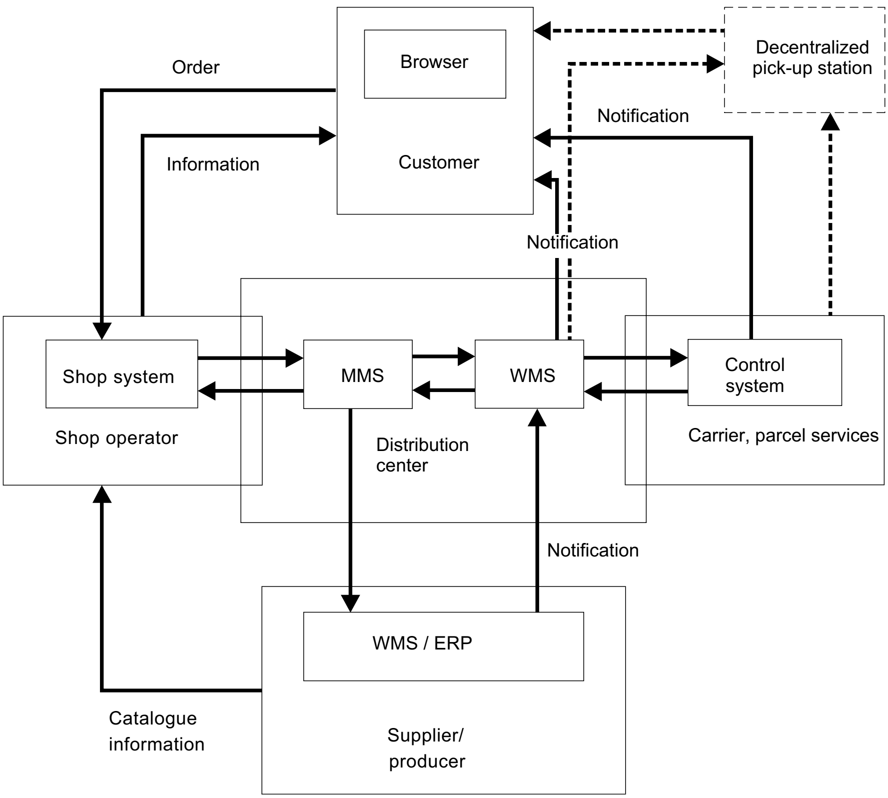


Figure 1.2. Basic information flows and system elements in e-Logistics

for the tracking of C.O.D.s and the management or returns. If the goods are not directly shipped to the customer but to a decentralized pick-up station where it can be collected by the customer, the delivery data are transmitted to the station automatically by the control system or manually by the supplier. Furthermore, a delivery can be announced by the WMS of the distribution center which in this case acts as supplier. The customer is informed by the pick-up station about collectable consignments. Finally, the control system informs the distribution center about finished delivery transactions.

1.4 Structure and goal of this book

After the term WMS and its context have been defined, Chap. 2 introduces the management of warehouse and distribution systems and shows the respective material flow processes. Here, a warehouse system is described by studying the typical material flows and processes. Main emphasis is given to the detailed and structured presentation of the processes, requirements and strategies in warehouse and distribution systems. Furthermore, the main order-picking principles are described as well as some special processes and the requirements on warehouse management systems. It is also demonstrated how the requirements of the technical structure of a warehouse management system can be fulfilled, i.e., how the tasks of the warehouse management and control in a WMS can be carried out.

Chapter 3 deals with the optimization of storage operations as well as scheduling and dispatching. This chapter mainly focusses on the optimization of the order dispatch.

Chapters 4, 5 and 6 give an overview over the basic technologies for the design and control of a warehouse and distribution system. These are the mechanical components like storage, conveying, sorting and distribution facilities described in Chapter 4, the general automation technologies in Chapter 5 as well as the principles of object identification in Chapter 6. Available technologies are described in the context of warehouse and distribution systems and only as far as they are of any relevance in this matter.

The general principles of information processing and communication in networks are studied in Chapter 7. This includes the basic elements of the operating system, communication in networks and the important matter of data security. The various aspects of the implementation of a warehouse management system from the requirements profile up to the putting into operation and the further information are discussed in Chapter 8.

Chapter 9 gives an example of the warehouse control and management. At first, the reader gets an overview over the basic architecture of warehouse management systems, i.e., WMS based on a data-base and distributed WMS. A warehouse management system is described based on this example, which includes all elementary parts of a warehouse and distribution system. As a bonus, a CD-ROM with a functioning system has been included.

2. Management of Warehouse Systems

Warehouse and distribution systems are tailor made for special requirements and pursuant to the variety of technical and organizational features they differ more or less. Nevertheless, the basic organizational processes are similar because these systems are part of a comprehensive material flow. This book intends to set a uniform basis for such systems and thus points out the major frameworks, basics and procedures, on the one hand, and the resulting requirements on the functionality of a WMS, on the other hand.

In an efficient warehouse management system, these frameworks and logistic entities are taken into consideration in order to achieve optimal processes. In the following, the basic warehouse processes and functions are described that are vital for efficient operation. Furthermore, the special role of order-picking as a labor-intensive and time-critical part of the internal process chain is studied. Finally, the typical characteristics and key values of the systems as well as the basic functions of a WMS are examined.

2.1 Logistic frameworks

2.1.1 Logistic principles

The term logistics

The term *logistics* describes the systematic approach towards the comprehensive optimization of flow systems, e.g., material flow systems, beyond single system boundaries. Depending on the alignment there are different definitions which are not dealt within this book. A lot of these definitions just describe the role of logistics in research and teaching, with regard to the planning, organization and control of such flow systems ([31, 40, 44, 62] et al.).

Since this book focusses on the material flow special emphasis is given to the so-called "6Rs" of logistics[1] The "6Rs" of logistics describe the logistic targets as the delivery of

[1] Sometimes also called 4Rs in correspondingly compressed contents.

– the right goods at the
– right time in the
– right quantity and the
– right quality at the
– right location at the
– right costs.

Despite this simplification this principle is widely accepted. In this context *right* means the fulfilment of customer requirements like *ordered, required, expected* and *minimum costs.*

Coordination (synchronization) of material flow and information flow
According to another recognized and relevant finding the logistic performance of a system in the sense of an optimal delivery service at minimum costs can only be achieved when material flow and information flow are well coordinated. In most cases this means an information flow which precedes the material flow and thus allows for the reservation of transport, storage or handling capacities. This chapter continuously points out the importance and realization of this paradigm.

Supply chain management (SCM)
As already described earlier, present material flows include a variety of factors which lead to multi-level steps or sub-systems, a *supply chain.* In the classical case where each element is self-sustained and just observes the behaviour of its customers (in the form of orders) and of its suppliers (in the form of delivery times) the slightest variations in the ordering behavior at the end of the supply chain may result in massive fluctuations of the ordering behavior (and thus of system stocks) at the beginning of the chain. This phenomenon, also known as bullwhip effect, was described by Forrester in 1950. But only today, thanks to efficient computer and communication systems are we able to master these complex correlations. The key factor is the processing and supply of important system information at supply chain participants (echelons). Known strategies in this context are *ECR*[2] and *CPFR*[3].

Although this book does not focus on supply chain management and the methods and tools to optimize these chains, the recording, processing and handling of information are of great importance also for the operation of warehouse systems. As a consequence, the decisive mechanisms and principles of the SCM will be highlighted at all relevant points.

[2] *Efficient Consumer Response* is the direct response to the inventory system (of the supplier).
[3] *Collaborative Planning, Forecasting and Replenishment* propagates the linking of all supply chain elements in order to coordinate order requirements and delivery bottlenecks early.

Outsourcing

In recent years there has been a clear trend towards the outsourcing of internal logistic services. The main reasons mentioned in this context are:

- Concentration of operational efforts on the core competencies, e.g., development and production of goods and release of management capacities.
- Reduction of logistic costs by economies-of-scale effects and higher system utilization by the licensed logistic providers.
- Balancing of seasonal work peaks or fluctuations which do not justify the generation of own capacities.
- Improved delivery services by increasing the customer presence or by reducing the delivery times. Many licensed logistic providers already have such a presence while it would be uneconomical to build up an own presence. One example is the implementation of an EDI solution with a secure customer connection for the generation of the required delivery notices.
- Circumvention of wage-related frameworks.
- Purchase/creation of logistic competencies.

There are many different forms of outsourcing. For example, a licensed logistics provider may take over and operate an existing warehouse. In other cases the stocks are transferred to the external warehouse of the licensed logistics provider where they are stored together with the stocks of other companies (multi-client system).

Anyway, this leads to new requirements on the operation of such outsourced warehouses. Since such services often are paid based on performed transactions it is of vital importance to make the activities measurable and transparent so that the customer is able to control the processes.

Solutions which combine different customers in one system are much more complicated. Here, goods of apparently the same value have to be subjected to different processes, e.g., during stock taking.

And last but not least it results in a variety of requirements on the warehouse management and the employed warehouse management systems. Depending on their specific application, they have to be highly transparent, general purpose systems which meet different requirements. No wonder that large logistic service providers typically have large manpower for information logistics. These requirements are described in detail in the following section.

E-Logistics

One area which has become more and more important in recent years is *e-Logistics*, which can be defined as

> an umbrella term for the planning, management and control of the goods, information and monetary flow along the complete supply chain via public and private networks (Internet, intranets), i.e., from the front-end over the customer's online order (business-to-consumer

(B2C), business-to-business (B2B)) up to the tracking and tracing and the customer services [14].

E-logistics is, thus, the connecting element of industrial activities in the age of the Internet. Experience has shown especially in this area that an efficient logistics and quick material flow systems decide whether a project succeeds or fails. The focus on the sales channel and a neglected fulfillment have at least accelerated the downfall of many E-commerce companies.

It becomes also clear, however, that classical distribution systems often fail to fulfill the requirements of e-logistics. The decisive factor is the highly flexible and quick handling of small but numerous customer orders in line with dynamically varying orders and article ranges.

CEP services and e-Commerce

For years the CEP services (courier, express and parcel services) have experienced dynamic growth. This success is closely connected to the already discussed ordering behaviour of the customers and the required reaction speed of the companies. Especially with regard to the future development of E-commerce forecasts predict further growth for this sector [14].

The changed distribution behaviour also has negative effects on the processes in the distribution centers where an increasing variety of small consignments has to be picked, recorded and booked within a short time. These requirements can only be fulfilled efficiently by consequently using suitable control and documentation systems.

2.1.2 Packaging and logistic units

To get a comprehensive overview all typical units which are in the widest sense *transformed* during these processes have to be known. According to [40] such a transformation process constitutes a change of the system status of goods with regard to time, location, quantity, composition or quality. The units are present in the warehouse systems in varying forms and combinations depending on the requests of the relevant sub-systems and operating means and on the customer requirements. Different terms have established which are above all used in logistic systems (although not stringently).

As a consequence, a study of material flows starts with the quantities and characteristics of the goods which have to be transformed in the respective system. The actors or decision makers at the sinks of such material flows put entirely different requirements on these units especially with regard to their composition. At the beginning of the process chain producers or manufacturers are anxious to produce goods in economical quantities (batches) and to transport them as efficiently as possible, i.e., with a minimum of work steps to the next customer. An important aspect in this context is the consolidation of goods flows and processes. The logistic and material flow systems of the producers generally are designed for the efficient handling of a limited number of articles in large quantities. For this reason the required economies of scale are achieved by means of a multi-level trade. These structures have changed only with the appearance of E-commerce where classical trading structures are sidestepped and where each producer is directly present for the final consumer.

Packaging

At the beginning of each goods and material flow there is the necessity to protect the goods from impacts during transport, transshipment and storage (TTS processes) to avoid a loss of quality. This is mainly achieved, on the one hand, by choosing a suitable packaging and, on the other hand, by building so-called unit loads.

In addition to its main function, i.e., to protect the goods during the TTS processes, packaging has to fulfill a lot of other requirements. In some cases, e.g., hazardous goods, it should also protect the workers and the environment. In order to facilitate efficient handling during these processes a packaging should allow for easier storage and transportation, e.g., in the form of uniform units which can be stacked or transported by conveyors. In the retail trade, packaging should have a high marketing effect. It may offer customers and consumers a possible *usability effect*, for example a recloseable seal or reuse for other purposes after the original contents have been removed. Finally, the *identification and information effect* plays an important role to identify the goods within the material flow (e.g., by printed barcodes) as well as for the display of goods in the shop shelves.

This variety of requirements, which partly counteract each other, generally cannot be fulfilled by one single packaging alone, but only by a well coordinated packaging system [40]. For this reason, different types of packaging have established for the different sections of the transport chain. Types of packaging are also classified according to the waste they produce:

Transport packaging: Protect the goods during the transport from the manufacturer to the distributor.

Sales packaging: Packaging which are used by the *final consumer* for the transport of goods or until their consumption.

Outer packaging: Additional packaging of sales packaging which allow for distribution of the goods in the form of self-service, protect the goods against theft or serve as marketing media.

Especially in German packaging regulation, these types of packaging are given special attention because this regulation obliges trade and manufacturers to take back used packaging from the customer. A more detailed description of packaging technologies and problems is given in [20, 40, 47].

Table 2.1. The goals of building units

Goal	Examples
Simplification and cost reduction	• Reducing transshipping processes • Avoiding unnecessary identification processes • Minimizing tests, measurements and counts
Standardization	• Adjustment to technical devices • Standardized interfaces to the goods pick-up • Standardized dimensions for use of universal devices
Exchangeability	• Facilitate pool operation • 1:1-exchange of identical food
Functions	• Improve stackability • Grant access

Logistic units

Although a packaged good represents a unit within any logistic system, it is not necessarily a *logistic unit*. Such units are built with the superior goal to combine single goods adequately so that the required logistic activities can be carried out in the most optimal way. The most common goals resulting in the building of logistic units are shown in Table 2.1.

In the ideal case only *one* logistic unit is chosen or built which passes the complete transport chain unchanged, i.e., it is the storage, transport, shipping and loading unit all in one. As said earlier, however, this goal is almost impossible to achieve. Especially in multi-level supply chains there is a trend towards even smaller units (Chapter 1, p. 2).

A logistic unit can be created by choosing a suitable form of packaging. One example is the consolidation of bottles with condensed milk into packages. Usually packages are the smallest handing unit. For transportation they are combined into *loading units.*

The *building of loading units* is the efficient consolidation of goods into larger, manageable units. These are handled by means of suitable *loading aids.* Depending on their form and/or function, these loading aids are classified as follows:

– bearing loading aids on which the goods are placed or stacked,
– enveloping loading aids which take up the goods and support them sidewise
– encasing loading aids enveloping the goods on all sides

For an improved product and transport protection the loading units are also protected with securing means [92]. The most commonly used procedures are shrinking, stretching and strapping [40].

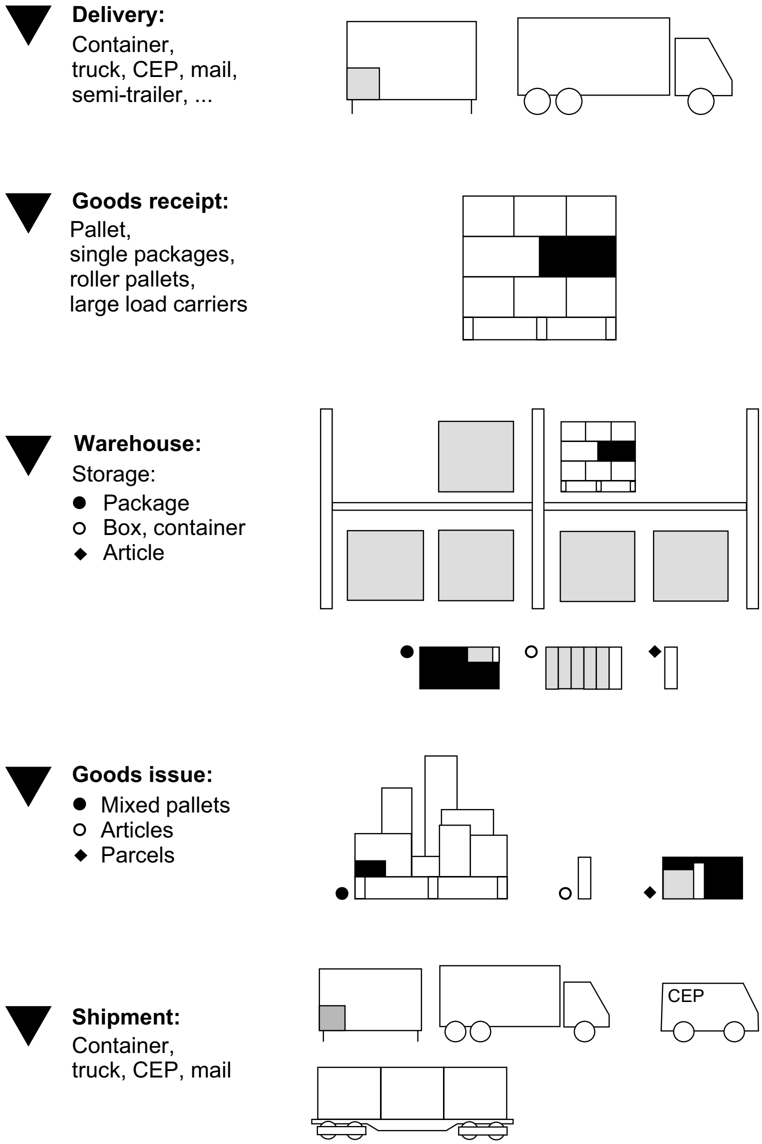


Figure 2.1. Change of the number of articles and loading units along an exemplary transport chain

As far as the material flow is concerned, the change between the different combinations of packaging and loading units is of interest, as is shown in Fig. 2.1.

Due to the special requirements of order-picking (cf. section 2.2.5) loading units have to be classified even further:

– *Storage units* are those units in which an article is stored in the warehouse (e.g., pallets, also supply units).
– *Supply units* are those units which are provided for retrieval (e.g., boxes or containers).
– *Retrieval units* are units of a certain article which are retrieved by the picker, if necessary with several single picks (e.g., packaging)
– *Picking units* include that quantity of articles or packaging which a picker can retrieve with one pick.
– *Collective units* are built when the picker collects the single items on a pick list.
– *Shipping units* represent the number of articles as ordered by the customer. The single shipping units are often built by means of loading aids (pallet, skeleton box, etc.).

2.2 Functions in warehouse systems

Although the requirements on a distribution center are quite diverse as are the processes, technical system designs and control functions each system still includes certain standardized processes. This is of vital importance since each distribution center is only a link in a subordinate supply chain. These typical basic processes as well as the resulting requirements on and functions of a WMS are described in the following.

2.2.1 Goods acceptance and receipt

After the goods have been ordered by the company's dispatcher the material flow starts with the delivery notification of the manufacturer or supplier.

Notification of goods receipt and delivery date Depending on the delivery status, a precise delivery date has been fixed especially in cases with a high number of deliveries and a low goods reception capacity. Such time schedules should help to avoid or reduce waiting times for trucks and, on the consignee's or loader's side, help to coordinate system loads and to avoid load peaks. This, however, depends on many parameters like the size and value of the consignment, the delivery distance and the general traffic situation, official working times (e.g., customs clearance) and, of course, on the location of the consignee and the ideal can hardly be achieved.

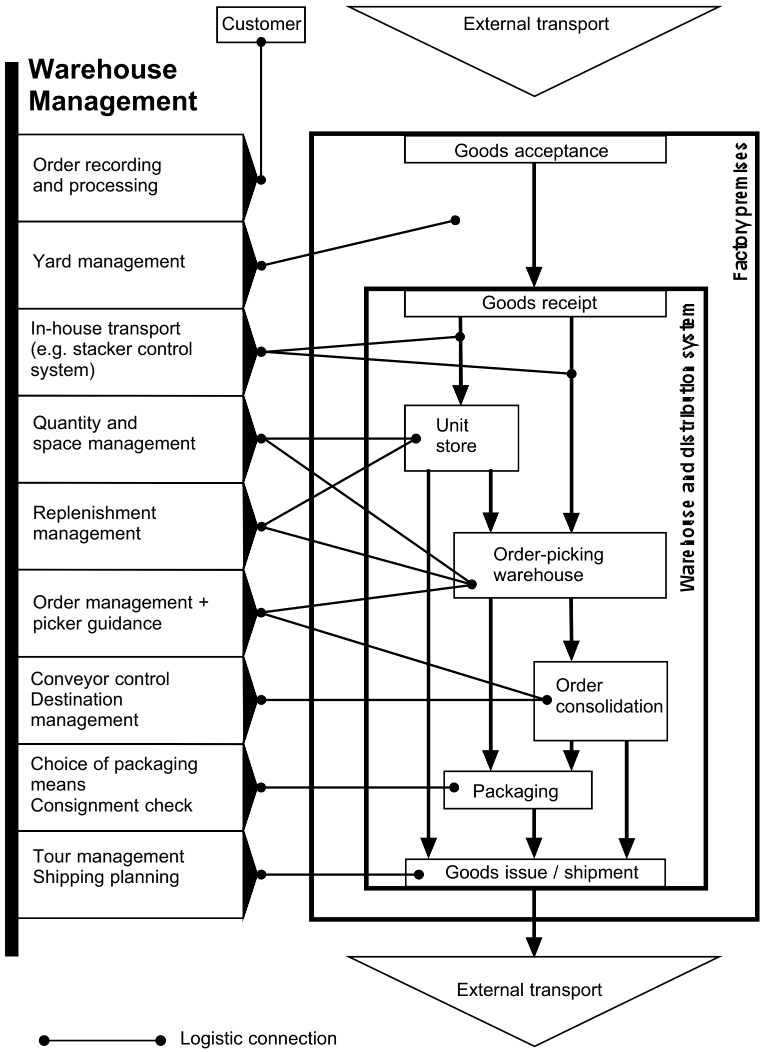


Figure 2.2. Basic elements of warehouse management systems and their role with regard to warehouse operations

Goods acceptance The goods acceptance is the first important step in the material flow of a warehouse. Based on the delivery note, the consignment is compared with the order and the bill of lading is compared with the notification. Then the notification data are temporarily entered into the inventory system. This process accelerates the goods receipt process considerably because in case of a positive check the data just have to be confirmed.

While in small warehouse systems goods acceptance and goods receipt may be in a physical vicinity due to the spacial separation in larger systems, arriving trucks have to be directed and assigned to the loading gates. The spacial separation mainly aims at a better control of the yard traffics. For this purpose, yard management systems (also stock management systems) are used to guarantee for a coordinated traffic on the premises and above all to minimize unnecessary searching and switching trips.

Goods receipt Based on the notification data, the goods receipt department can also be informed about the pending delivery. This is of special advantage in case of larger deliveries (with regard to the reception system). This includes, for example, the planning and reservation of buffer space, the choice of a suitable terminal (e.g., loading gates or bays) or the print of in-house labels for the internal identification of goods. Cross-company labelling systems similar to EAN128 used in trade or Odette in the automotive industry are advisable although they are not the state of the art in most warehouses. One reason, among others, is the fact that a material flow control puts other requirements on a labelling and identification system than the trade (cf. Chapter 6).

Incoming goods inspection In addition to the logical comparison of ordered and incoming goods, the consignment is submitted to a physical check. This includes the inspection of all goods with regard to type and quantity is generally performed by the unloading staff. The condition of some goods is carefully checked by the quality assurance according to the company rules. These inspections may be a simple visual test, laboratory tests of samples or a complete full-scale control. Faulty goods are marked with a blocking indicator and moved into special areas or stored under consideration of the blocking indicator (section 2.2, p. 27).

In case of new articles the article master data have possibly to be completed (cf. Table 2.13). The careful collection and maintenance of these data are vital for a number of control and optimization functions further down the material flow process. Exact weights have to be known for order-picking tests so that it can be checked if customer orders are complete. Special automatic devices like picking robots need these information, e.g., to verify a successful gripping process. For this reason, all new articles or articles still unknown to the system have to be weighed at the incoming goods inspection.

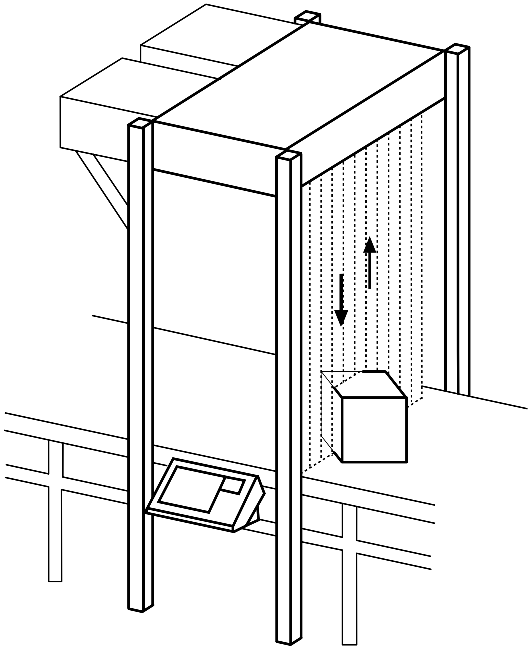


Figure 2.3. Automatic determination of unit dimensions (according to [59])

The dimensions of an article are also of importance to optimize the volume utilization along the material flow. The racks in a high-bay warehouse can be adapted to the article range by changing the shelf heights and thus to optimize the utilization of the storage capacity. Furthermore, according to the volume of a picking order, the optimal shipping containers can be found and thus shipping costs can be minimized.

The article dimensions are determined by different methods. In addition to measuring or entering the exact article measurements into the warehouse management system some users are content with a rough classification into different size groups. This method only makes sense when the articles are measured manually. For this purpose often gages are used which bear marks for the respective size categories so that it can easily be identified by the worker.

Furthermore, there are also automatic measuring devices which scan the article and transmit the measured dimensions directly to the warehouse management system (cf. Fig. 2.3). Then other characteristics of the article are recorded when the article master data are not yet available.

In order to secure high-quality goods, it may be necessary to record the serial numbers as well. The continuous control and traceability requires functions beyond the goods descriptions since not the article but the special loading unit and its movements in the warehouse should be documented.

Furthermore, the information may have been stored for a while after the unit has passed the material flow. The same applies to the tracking and tracing of production batches.

In order to avoid the perishing of goods with a limited lifetime, it may also be necessary to record the expiry date or the remaining life. These parameters can serve as basis for retrieval priorities or stock transfers.

Building of loading units In many warehouse and material flow systems special loading aids are used for handling or security reasons. These are, among others, tray storage or rack systems with standardized containers. Incoming goods are consolidated into volume and quantity-optimized units with the aim to minimize shipping and transport costs. Here, the goods have to be refilled in company-specific containers and consumption units to fit the material flow system. The same applies to units on standard pallets (e.g., Euro or industrial pallets). Often, the quality of the delivered units is questionable with regard to the downstream automatic systems like a high-bay warehouse. Damaged pallets, for example, may exceed the tolerances and thus get stuck at certain sections of the conveyor belt. For this reason, such units often are restored on tailor-made storage units (undamaged high-quality standardized pallets) even if this means a worse utilization of the storage space. This method allows for the permanent tagging of pallets for a better control. If highly dynamic automatic storage technologies are used incoming loading units may have to be secured additionally.

If the volume of the goods to be stored is much smaller than the available minimum storage volume (e.g., the shelf of a rack) usually *mixed pallets* are built to improve the utilization of storage space. This is of special importance for the controlling warehouse management system. The additional storage of articles with widely differing characteristics which call for the precise management of article dimensions and available storage capacities is not the only problem. When goods are retrieved it has to be taken care that only a certain article is taken from a mixed pallet. It has to be avoided that the number of partially loaded pallets becomes to large. This may require a volume control and restorage, repackaging or densification (cf. section 2.3.2, p. 50).

A particularity is the *handling of returns* (return of goods by the customer) in the mail-order trade. In classical mail-order trade, the rate of returns may be 30% of deliveries or more. Since the quality of the single articles is not known in advance all returns have to be checked carefully. Thus, returns are single incoming parts which have to be checked separately, cleaned, repacked and labelled. This requires a considerable amount of human labour. Depending on the used strategy the returned goods are stored at the current single-item partial pallet or stored in special locations together with other returned goods with a corresponding retrieval priority.

When all tests have been passed successfully, the articles are suitable for storage and are inventoried by the system.

2.2.2 Storage

In case of a manual storage system (e.g., stacker-operated block warehouse), the process is continuous because the loading unit is taken up and then directly moved to its final storage bin where it is stored or transferred. In larger automatic systems, however, the process is performed stepwise up to the final storage in the storage bin. For this reason, the overall process is described below. In a manual warehouse it only consists of specific partial steps.

Distribution to storage areas In a first step the back orders have to be checked (especially in manual systems), so that articles may be transported directly to the place of consumption or into the shipping area. This process is also called *split-lot storage*. Another term used simultaneously is *Cross Docking* which more precisely describes an extensive concept (cf. page 59) and thus is not the right term here.

Then the articles are moved to the storage areas. For this purpose, the transport targets have to be determined in the warehouse management system. Above all in large systems, there are large optimization potentials because of the long transport distances and large quantities, especially when manual unsteady conveyors (e.g., stackers or tractors) are used. The in-house transports can be optimized with different methods. On the organizational side, there is the collection and presorting of tours what requires sufficient buffers and room for action. This method may result in different handling times for the goods to be transported. This can be avoided by using an automatic transport control system which controls the transport of articles into the different storage areas by means of deterministic strategies and rules. For this purpose, the stacker cranes can be equipped with a control system which provides an optimal order sequence (cf. section 2.3.3, p. 50 and Chapter 3).

An important aspect is the *transparency* of the material flow, above all in larger material flow systems. For example, when a transport unit is moved to a wrong location this typically requires a physical search. A suitable measure to trace the goods and identify possible errors is to treat the conveyor like a *virtual storage bin* and to rebook the goods to these conveyors during transports. This results in a continuous documentation chain and single units can easily be tracked and traced.

Identification If the storage unit has not yet been identified during the goods receipt check this *identity control* is now carried out at the so-called *identification point*. This is often done in the prestorage zone of automatic warehouse systems like high-bay warehouses. For this purpose, it is checked whether the article and quantity correspond with the loading unit and if the master data is available. At the same time the material flow is synchronized with the information flow. When automatic warehouse systems are used this

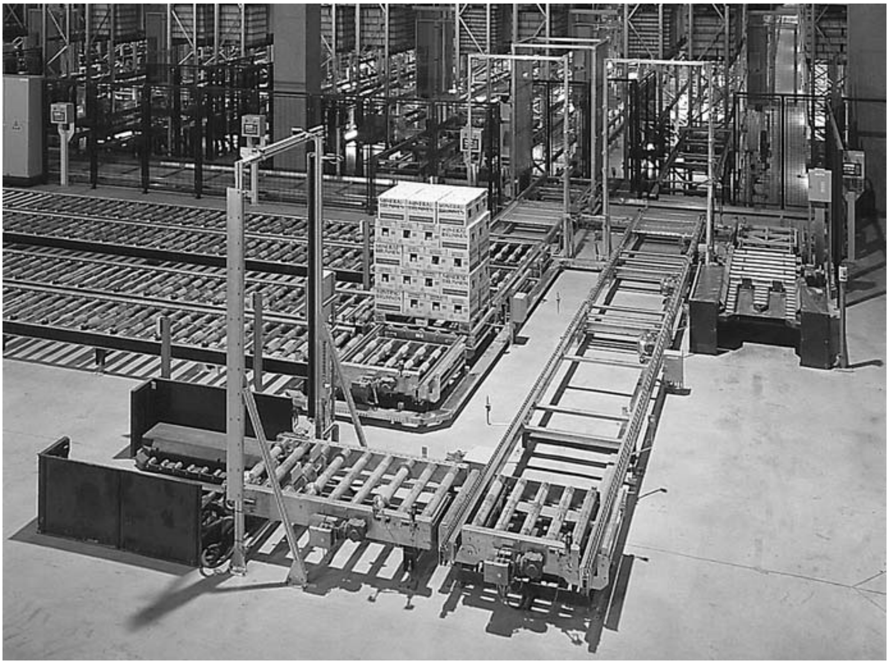


Figure 2.4. Identification and contour check at the identification point in front of a high-bay warehouse
[Photo: STÖCKLIN SIEMAG]

happens at interfaces between the manual and the automatic system. From now on the order cannot be changed.

At the same time the *stackability* has to be proven. This includes first of all the *contour check* (check of the dimensions of the goods to be stored) and, if necessary, a weight control, especially in systems with non-uniform shelf capacities (Fig. 2.4). This may be followed by a physical stock taking.

Assignment of the storage bin and put-away The first step in a storage process is the determination of the storage bin. This may be done according to quite a variety of criteria which result from the physical requirements of the goods to be stored, the operational and technical warehouse operation as well as from security and legal requirements (cf. Table 2.2).

The requirements concerning the physical dimensions and weights of goods to be stored arise from the construction of economical shelves where specific dimensions and loads have to be determined from the current and future range of articles. When the article range can be segmented in such a way (sufficient classification) the allowed loads usually are reduced for each shelf in upward direction or corresponding load areas are built. In manual

Table 2.2. Some important parameters for the assignment of storage bins

Parameter	Requirements
Technical requirements	→Observance of allowed shelf and field loads →Even shelf load and avoidance of one-sided loads above all in dynamical warehouse systems →Optimal utilization of shelf volume
Operational optimization	→Minimizing driving and transport ways →Maximizing turnover rate →Maximum utilization of the storage capacity →High availability, i.e., access security also in case of failure of single transport or operating facilities →Quick detection and identification of goods in manual systems
Security and legal regulations	→Observance of consolidation restraint (storage of hazardous goods) →Separate storage (Food) →Batch grouping

warehouses, lighter goods are stored in the top shelves for ergometric reasons.

In some dynamic warehouse systems like the horizontal flow racks (cf. p. 104), a one-sided load has to be avoided for functional reasons. Here, the system control has to distribute the goods evenly to optimally utilize the available storage volume. For this reason, shelf heights should be graded to take up goods of different heights.

A variety of strategies can be used to optimize the operating process of a warehouse system which is partly incompatible. The most commonly used strategies and their compatibility are shown in Table 2.3. In this context the shortest way strategy is a generally applicable secondary strategy and mentioned for the sake of completeness.

Of course, security and legal regulations have priority, above all with regard to hazardous goods and food. Here, a large variety of special regulations have to be observed, which in the following are only described with regard to their effects onto warehouse management. The *storage monitoring* is also an integral part of a modern warehouse management which terminates the storage process. For this purpose, a feedback is given showing the storage bin and time. While in automatic warehouses this step is carried out automatically in manual warehouse systems (with a computer-aided control, e.g., stacker control system), it has to be verified by an operator, e.g., by directly

Table 2.3. Operational storage bin assignment strategies

Description	Strategy	Objective	Compatible			
Fixed storage bin	Fixed assignment of a storage bin for a certain article	Access security for choice of the management system Quick access in manual order-picking systems with practice (reduced search times)	•	•		
Random storage (chaotic storage)	Article can be stored randomly at any free storage bin	Maximum utilization of storage capacity			•	•
Zoning	Choice of a storage bin according to turnaround of the article	Increased handling performance by minimized medium path length	•		•	
Cross distribution	Storage of several units of an article over several aisles	Availability of the article in case of a rack feeder breakdown Increased handling performance by parallelization	•	•	•	•
Pick/parts families *Clustering*	Adjacent storage bins for frequently consolidated articles	Increased handling or picking performance by minimized connection ways		•		•
Shortest driveway	Approach of storage bin on shortest way	Increased handling performance by minimized connection ways	•	•	•	•
Pre-buffers	At peak times storage in front storage area	Avoidance of tailback by increased throughput			•	•

recording the goods and storage bin. In established systems the next order is released only after this step has been validated.

2.2.3 Retrieval / picking

Retrieval orders are managed over a longer or shorter period of time according to the situation. In any case, the present orders have to be checked with regard to their feasibility. The quantities and/or units to be retrieved are blocked to avoid shortages at the fixed retrieval time (cf. section 2.2, p. 27).

Table 2.4. Retrieval strategies

Description	Strategy	Objective
FIFO (First-In-First-Out)	Retrieval of the first stored loading unit of an article	Avoidance of obsolescence and expiry of single loading units of an article
LIFO (Last-In-First-Out)	Retrieval of the last stored loading unit of an article	Avoidance of stock transfers in case of certain storage techniques (block warehouse)
Quantity adaption	Retrieval of full and broken loading units according to the order volume	Improved handling performance by minimized restorage
Partial unit preference	General prioritization of partial loading units	Improved utilization of the storage capacity
Shortest driveway	Retrieval of loading units of an article on the shortest connection ways	Improved handling performance by minimized driveways
Minimized aisle swaps	Sorting the retrieval sequence according to single warehouse aisles	Minimized shifting for curve-going rack feeders or flow racks
Tour-related	Sequencing of retrievals according to the tour schedule of a downstream transport means	Reduced switching and reloading
Time-phased	Scheduling of the retrieval according to the presumable time of demand	Reduced switching and reloading
Advancement	Restorage of soon to be retrieved units near the transfer point	Reduced reaction times by increasing the handling performance at the time of demand

The dispatching and performance of the retrieval are carried out by means of certain retrieval strategies considering different objectives (cf. Table 2.4) Some strategies (time-phased or tour-optimized retrievals) can only be used in line with a corresponding dispatching.

Furthermore, the warehouse management system has to monitor the initiated retrieval process and give feedback about its correct completion. When the goods have been retrieved the storage bin is released, the inventory is updated by reducing the stock by the retrieved quantity and corresponding reservations are cancelled.

To be able to continuously track the material flow, the loading unit, respectively the inventory, should be booked on the next recipient or the transport means (which will logically be treated like a storage bin).

2.2.4 Consolidation point

The consolidation point is a location where handled units are identified at key material flow positions. Such a measuring point is of great importance for demonstrating the production progress in the warehouse above all in systems submitted to stochastic influences, e.g., manual work processes.

On the one hand, the destination and actual data are compared to check the consistency of the material and information flow. Here, the order status is generally updated, i.e., the status of the retrieval order. The control point sets the basis for a dynamic capacity utilization, i.e., for the utilization of the following areas according to the work load. On the other hand, decisions concerning the material flow are made, like choosing the transport targets for finished units.

2.2.5 Order-picking

The consolidation of a customized quantity of one or several articles is called *order-picking*. Order-picking describes the building of customized units, i.e., the retrieval of items from a larger unit of single articles and their consolidation and provision for shipment.

Order-picking is a labour-intensive and expensive part of a warehouse and distribution center. It is, therefore, given special attention during the planning and operation of such systems.

In modern order-picking all kinds of conveyors and storage equipments are used. Because of the close relationship between technical facilities, process and organizational structure and information management the design and operation of order-picking systems is a highly complex task. Basic functions and standardized processes have been defined to improve the structuring of these processes and tasks which should allow for a systematic planning and organization of an order-picking system [38, 40, 88, 89]. A traditional order-picking system is divided into the areas:

- material flow system
- organization
- information flow

Material flow system

During the design of the material flow of an order-picking system, it must first be determined how the pickers and articles can be brought together most efficiently and in which form single retrieval units resp. collection or picking units (cf. section 2.1.2, p. 20) should be transported further.

Just in some cases (e.g., automatic picking facilities like shaft picking systems) there is no direct movement because the separating process — similar to a cigarette automat — is carried out in the machine itself and the goods are transported to the collection point from there. According to [40], the physical order-picking process consists of the following basic material flow functions

– Movement of goods for pick-up
– Supply
– Movement of picker to pick-up station
– Retrieval of goods by picker
– Transport of retrieval unit for transfer
– Transfer of retrieval unit
– Transport of picked unit to transfer station
– Transfer of picking unit
– Return of broken loading unit

Based on the morphological kit shown in Table 2.5 structures and solutions for order-picking systems can be found and described by combining the single elements. It has to be taken into consideration that in order to consolidate picker and goods either the picker or the unit has to move. Furthermore, either the retrieval unit or the picking unit has to be transported to finish the picking process. The "picker" may either be a human or a machine (e.g., a picking robot).

In this context special attention should be given to some classifications. Above all terms like "statical" and "dynamical" are used differently in practice and literature. The classical definition which is also used with regard to warehousing means that at a statical supply a unit remains at the same place from storage until retrieval (cf. e.g., [30]), i.e., that it awaits its pick-up, for example in a rack. For a dynamic supply the unit of the desired article has to be brought to the pick-up point and returned after the retrieval.

In more recent publications these terms describe the retrieval process itself. It means that at a statical supply the parts to be gripped are statical while at a dynamical supply they are *in motion*.

Another aspect of this problem is the classification into *centralized* and *decentralized* supply. *Centralized supply* means the supply and retrieval at a fixed location or at least in a limited area (e.g., 2–3 adjacent pallet transfer points or picking out of horizontal rotary racks (cf. section 4.1.3, p. 104)). The units are provided in sequence at this central point and the picker has access only to these units. At a *decentralized supply* the articles are retrieved at different points to be approached by the picker.

In literature the statical or decentralized supply is equated with the principle *man-to-goods* and the dynamical or centralize supply with the principle *goods-to-man*.

Given the fact that the classification into "centralized–decentralized" does not cover all aspects of this problem and that picking methods where the

Table 2.5. Basic functions of a material flow system according to [40]

<table>
<tr><th>Basic functions material flow</th><th colspan="5">Possible implementations</th></tr>
<tr><td rowspan="3">Moving goods to pick-up station</td><td rowspan="3">no movement</td><td colspan="4">Movement</td></tr>
<tr><td>1-dimensional</td><td colspan="2">2-dimensional</td><td>3-dimensional</td></tr>
<tr><td>manual</td><td colspan="2">mechanical</td><td>automatic</td></tr>
<tr><td rowspan="3">Provision</td><td colspan="2">statical</td><td colspan="3">dynamical</td></tr>
<tr><td colspan="2">centralized</td><td colspan="3">decentralized</td></tr>
<tr><td>sorted</td><td colspan="2">partly sorted</td><td colspan="2">unsorted</td></tr>
<tr><td rowspan="3">Picker moves to pick-up station</td><td rowspan="3">no locomotion</td><td colspan="4">Locomotion</td></tr>
<tr><td>1-dimensional</td><td colspan="2">2-dimensional</td><td>3-dimensional</td></tr>
<tr><td>manual</td><td colspan="2">mechanical</td><td>automatic</td></tr>
<tr><td rowspan="2">Picker retrieves goods</td><td>manual</td><td colspan="2">mechanical</td><td colspan="2">automatic</td></tr>
<tr><td colspan="2">single item</td><td colspan="3">collective item</td></tr>
<tr><td rowspan="4">Transport of pick-up unit to transfer station</td><td rowspan="4">no transport</td><td colspan="4">Transport</td></tr>
<tr><td colspan="2">picker</td><td colspan="2">conveyor</td></tr>
<tr><td>1-dimensional</td><td colspan="2">2-dimensional</td><td>3-dimensional</td></tr>
<tr><td>manual</td><td colspan="2">mechanical</td><td>automatic</td></tr>
<tr><td rowspan="3">Transfer of pick-up unit</td><td colspan="2">statical</td><td colspan="3">dynamical</td></tr>
<tr><td colspan="2">centralized</td><td colspan="3">decentralized</td></tr>
<tr><td>sorted</td><td colspan="2">partly sorted</td><td colspan="2">unsorted</td></tr>
<tr><td rowspan="4">Transport of picking unit to transfer station</td><td rowspan="4">no transport</td><td colspan="4">Transport</td></tr>
<tr><td colspan="2">picker</td><td colspan="2">conveyor</td></tr>
<tr><td>1-dimensional</td><td colspan="2">2-dimensional</td><td>3-dimensional</td></tr>
<tr><td>manual</td><td colspan="2">mechanical</td><td>automatic</td></tr>
<tr><td rowspan="3">Transfer of picking unit</td><td colspan="2">statical</td><td colspan="3">dynamical</td></tr>
<tr><td colspan="2">centralized</td><td colspan="3">decentralized</td></tr>
<tr><td>sorted</td><td colspan="2">partly sorted</td><td colspan="2">unsorted</td></tr>
<tr><td rowspan="3">Return of partial unit load</td><td rowspan="3">no return transport</td><td colspan="2">Return to warehouse</td><td colspan="2">Return to partial pallet warehouse</td></tr>
<tr><td>1-dimensional</td><td colspan="2">2-dimensional</td><td>3-dimensional</td></tr>
<tr><td>manual</td><td colspan="2">mechanical</td><td>automatic</td></tr>
</table>

retrieval of an article in motion is practically unknown the following definition is used:

> The differentiation into *statical* and *dynamical* supply defines whether the unit has to be transported by a conveyor for the purpose of retrieval.

And:

> The differentiation in *centralized* and *decentralized* supply defines the retrieval site. At a centralized supply the articles are retrieved at a fixed location while at a decentralized supply they are retrieved at different locations.

These classifications are described in Table 2.6, which shows that the statical supply is not equivalent to the movement of picker to goods (*man-to-goods* (MtG) because in a picking pool the retrieval can also be centralized. The dynamical supply may also required a movement of the picker. In these cases the term *goods-to-man* (GtM) is no longer precise and it cannot clearly be differentiated between GtM or MtG.

During the transfer of picked units, these classifications sometimes get another meaning. As for the transfer of retrieval or picking units the classification into statical and dynamical transfer refers to the transferring conveyor or the collection bin. If the conveyor is in motion (continuous conveyor system), the transfer is dynamical. If the unit is transferred to a stationary collection bin, however, the transfer is statical. The classification into a centralized or decentralized transfer is analogous to the retrieval: if, for example, a collection bin is carried along the retrieval units are transferred at different locations, i.e., decentralized, while a transfer at fixed locations is considered as centralized. Different examples are given in Table 2.7.

Further classifications are made with regard to the kind of movement, the kind of retrieval and the arrangement of already provided or transferred goods. The picking movement may be one-, two- or three-dimensional. When the picker moves on ground level along a shelf his movement is one-dimensional. A two-dimensional movement can be made, for example, by a rack feeder or a stacker crane, while three-dimensional movements are performed by cranes. Most retrievals are carried out manually, while heavy or bulky goods are retrieved with the help of mechanical devices. In case of suitable article and order volumes, the picking may also be automated by means of picking robots or shaft picking systems, e.g., in the pharmaceutical industry. The arrangement of the units during the supply and transfer is also of great importance with regard to automation. The picking can be automated the more easily and efficiently the better the units are arranged.

While basic parameters and characteristics can be determined the decision about which system is the best for a certain application practically depends on the system design.

Table 2.6. Example of a supply

Description	Statical	Dynamical
Decentralized	*Shelf rack system* Articles are provided in a shelf rack. The picker moves along the shelves and retrieves single units according to the order data. Only those units are approached for which a corresponding demand exists. This process is also called the *man-to-goods* principle.	*Shelf front at ASPW* The supply unit is stored, for example, in an automatic small parts warehouse (ASPW). The articles are picked from shelves on floor level at the side of the ASPW. The supply units are provided dynamically in the bottom shelves. But different bins are approached so that like in a shelf rack system the picker moves along the shelves.
Centralized	*Picking pool* The shelves are arranged (mostly in U-shape) and the picker stands in the center where all articles are within his range. Since there are no trips the picker achieves a very high picking performance (up to 1000 parts/h). The method is restricted to the picking of a limited number of small articles.	*Prestorage zone of a high-bay warehouse* The units are in an automatic high-bay or small parts warehouse and have to be brought to a centralized transfer point for retrieval. After they have been retrieved from the shelf the storage units are brought, mostly by a steady conveyor, to the picking station and then restored. Such an arrangement is also called a *picking U* and the process is defined the *goods-to-man* principle.

Organizational forms
The organization of the picking system, i.e., the structure and control of the processes within the system, has a large influence on its efficiency. Typically, the organization is differentiated into the *structural organization*, i.e., the arrangement of the storage areas and the *process organization*, i.e., the performance of picking processes.

Structural organization The structural organization defines the suitable structure for an order-picking system. Main emphasis should be given to the question which supply system is the best for which article. This decision

Table 2.7. Example of the transfer of a pick-up unit

Description	Statical	Dynamical
Decentralized	*Pick-to-box* The picker places the units into an entrained container (pick tray) He moves with the tray between the retrieval points. If the tray is at the same time the shipping unit for the customer this principle is called *pick&pack*.	*Pick-to-belt* Directly after the retrieval the picker places the unit on a mostly driven conveyor which runs parallel to the shelf front. Then he moves to the next pick-up point.
Centralized	*Goods-to-man/ picking-U* The units retrieved at the pick-up point are transferred to a collection unit (pallet or container) where they are stacked.	*Goods-to-man / paternoster rack with roller conveyor* The units retrieved from the paternoster shelf are transferred to a belt conveyor installed in front of the shelf. The picker does not move.

requires a careful analysis of the product range and the order structure which results in varying requirements on the capacity, performance and features of the supply system. These requirements are determined by:

– Volumes, weights and dimensions of the supply units
– Transshipment or gripping frequency per article
– Average quantity of each article to be retrieved per time unit, grip
– Frequent combinations of single articles
– Security requirements (high-quality goods)
– Temperature and security requirements

Since each supply system has its own special qualifications (cf. Chapter 4), it could make sense to use different systems. For this reason dedicated *zones* are built for different kinds of articles. But an optimization can be achieved also within a supply system by means of a logical zoning (cf. Table 2.3).

Process organization The productivity of a picker is influenced by:

– Basic time (e.g., order acceptance, sorting of bills, take-up of picking containers, transfer of goods or picking containers, transfer or final preparation of documents)
– Gripping time (approach, take-up, transport and placement of retrieved units)
– Idle time (e.g., reading, ripping of packages, searching and identification, control and reaction)

– Way time (movement, i.e., driving or walking) of the picker between acceptance point — pick-up point — transfer point)

The sum of these time slots is called picking time or *average picking time* as far as the averaged time shares are taken into consideration. On the one hand, it is determined by the order structure and mainly by the average number of lines (order items) per job. On the other hand, it strongly depends on the system structure and the organization of the units to be retrieved and the pick-up points. While the basic and idle times can be influenced by choosing a suitable information system (cf. p. 39) the process organization focusses on the gripping and way times.

In the most simple case, the job is handled by just *one* picker who carries out this job completely before he starts the next job. This principle is described as a *simple, job-wise picking*. This process may be quite reasonable with regard to the man-to-goods principle where the average order volume utilizes the picker's transport capacity. Another advantage is that it requires less preparation. In the minimum case the arriving order is directly used as picking list. However, the picker has to cover long distances because the picking order is determined by the order so that this principle is suitable only for small systems.

The first possibility for optimization is to collect and process several customer orders at the same time what is called *order-parallel picking* or *sorting during picking*. By increasing the density of pick-up points, the average way time per job can be reduced. The picker has to be guided through the system in such a way that he automatically reaches the next pick-up point and thus long idle times (identification of the nearest article) or returns are avoided.

In large systems it does not make sense to let a picker pass the whole system for one order. In this case the picker would have to know all areas and cover large distances. Furthermore, this method would lead to an increased traffic and uncoordinated processes. For this reason, the areas are divided into small sections where one or just a few pickers handle a part of the customer order. When the articles stored in the respective zone have been picked the collection bins are transferred to the next section. This method is called order-picking in *serial zones*. It is of special advantage that single zones can be bridged with conveyor systems if no articles have to be retrieved from there.

As an alternative, the customer orders can be divided into partial jobs which are assigned simultaneously to the single zones where they are collected. This reduces the order lead-time. This method is called parallel or more exactly *zone parallel order-picking*. On the other hand, however, there is the expenditure for the preparation of orders and the consolidation of sub-orders. While orders can be executed mostly computer-aided buffers, collection or distribution systems may be required to consolidate the sub-orders (cf. section 2.2.6, p. 44).

All these methods have in common that the relation of the article to the respective order is visible throughout the process. These methods are called *single level* since the articles are retrieved and assigned to a customer order in one step. The way times can only be reduced to a limited extend by means of an order-parallel picking (simultaneous approach of similar or adjacent articles or pick-up units) because the order has directly to be divided into sub-orders. In article-oriented picking, on the other hand, the retrieval and consolidation processes are separated and carried out in two steps . With this method all identical articles in a larger order volume can be picked within one picking process, i.e., the pick-up unit has to be approached or transported to the pick-up station only once. The way times as well as the gripping times can be reduced drastically.

This method requires the consolidation of several orders into so-called *batches*, thus the term *batch picking*. After all units in the batch have been collected the single pick-up units are assigned to the customer orders. This is done by special conveying machinery, so-called *sorting and distribution facilities* or *sorters*. The batch picking requires a relatively high system expenditure for the preparation of orders, the transport of pick-up units and the distribution of orders so that it is uneconomical for small systems with a low order volume.

The article-oriented picking (two-tier picking, batch picking) renders a high performance but also requires the following:

– Good transportability of pick-up units with similar dimensions and handling characteristics
– Computer-aided order execution and consolidation for the sorting of pick-up units and assignment to customer orders.
– Sufficient possibility to consolidate received orders, i.e., sufficient order quantities for batches with a similar priority.

Operational organization / control strategies The operation of an order-picking system requires different rules, strategies and flexible behavioral patterns to meet the daily varying system requirements. As a consequence, the operational organization of a picking system is represented by a series of rules which are adapted to the special system requirements and can either be part of the warehouse management system or a statistically established rule. These are activities like:

– Input or handling of fixed dates and express orders
– Order input depending on the currently available picking capacity, the current work progress or the system status
– Assignment of pickers to zones or activities (resource management)
– Initiating supply

Principally, all measures are taken which are treated under operational organization in Chapter 3 (see there).

Information processing

Information processing means the collection, preparation and processing of all information which are necessary for the order-picking. These are [40, 88]

- The collection of order data under consideration of customer services and the capacity of the recording system,
- the processing of orders in a format which is compatible with the type of picking organization,
- the guidance of the picker by assigning pick-up stations and quantities and
- the process control.

Order-data collection Customer orders have to be collected securely and efficiently while at the same time there has to be a high-quality customer service. This leads to the following requirements on the data collection:

- A selection of possible order placements (by phone, fax, data transmission, Internet)
- Time of order placement according to the customer's wish (=optional time)
- Direct information about availability of articles and a possible delivery date
- Possible consultation with customers
- If necessary, consideration of special requests

On the other hand, there is the goal of a cost efficient and error-free order acceptance. Peaks have to be avoided for capacity reasons. For this reason, often service providers (call-centers) are used for critical products. A solution for business customers are fixed request dates for regular orders. Because of the good service and high availability Internet-based systems have more and more influence. The current availability of articles is of critical importance for all collection channels. This requires a precise and quick stocktaking. The collection of order data is a typical function of an enterprise resource planning system (ERP) but not of a WMS. This function refers, however, to the current and realtime inventory management.

Order execution The recorded order data is unsuitable for an efficient order-picking, except in some very small systems. Depending on the chosen picking method the following activities arise:

- Completion of orders with the picking-relevant data (storage bin, article number, collection bin)
- Sorting of items in the order of their arrangement in the shelf
- Division of orders into sub-orders which are executed in different zones
- Consolidation of orders into batches which have to be handled jointly
- Filtering of orders with the same priority, type of shipment or destination
- Filtering of orders with different treatment (e.g., single-item orders)

Information management

Picker routing In most order-picking systems, the operator, often supported by technology, works as a picker. The following section focusses on how to guide a human picker through the picking system. Automatic picking systems, on the other hand, mainly focus on certain article ranges where their special features have the optimal effect. Anyhow, automatic systems require specialized, system-specific controls.

It is the main task of the picker routing to transmit relevant pick-up information to maximize the picking performance and to minimize possible picking errors. The transmission of pick-up information to the picker can be divided into voucher-bound and voucherless processes.

Faulty pickings not only have a negative effect on the customer's trust of the logistic abilities of the supplier, but also cause considerable financial loss and thus are a critical system parameter. In order to minimize errors, controls are performed at different stages of the picking process. Suitable methods largely depend on the used picker routing system.

In addition to avoiding picking errors these measures should help to record the system status and finished orders to enter the next orders. For this reason the control is also called *confirmation* of the order or pick-up unit.

Picker guidance with picklists The voucher-bound method is the classical form of a picker guidance. The picker gets a sheet of paper containing the picking information. Principally, this picklist can be used for all picking methods. In automatically supported systems, however (automatically controlled goods-to-man systems), it makes not much sense because the information which are necessary to control the plant (e.g., horizontal or vertical rotary racks) are already available in an electronic form.

What is decisive for the picking performance is the order of pick-up information on the list. This should correspond to the order of articles on the shelves. In larger systems where the pick-up units are arranged optimally with regard to volume and way, the picklist has to be rearranged according to the customer order which typically is arranged according to customer or article number or alphabetically. This can only be achieved by computer-aided management systems. In zone-parallel picking the original list has to be divided into several partial lists.

The picklist has the advantage that it is easy to prepare and make out and in ideal cases also additional functions can be carried out. A common form of picklists are labels by which the pick-up units can be tagged with various information (e.g., prices). This helps to save work steps.

A self-control of the picker is often achieved by checking off the single items on the picklist and acknowledging the list, e.g., by signing the finished picklist. A further control is only possible by the following 100% control of the complete customer order.

Voucherless guidance and control of the picking process Disadvantages of the picklist are the long inefficient time spent to identify the next pick-up position and handling of the list but above all the enormous inflexibility. The basic times which are necessary to prepare and print the lists cannot be reduced and changes on short notice are difficult to add. This becomes evident in case of shortages which have to be reported and handled manually. For these reasons, the picklist system requires that the picking system react quickly to a changing system status (these also include dynamical batch control processes in the two-tier picking).

Table 2.8. Voucherless picker routing

Description	Function
Mobile terminal	The picker receives the pick-up information online (via infrared or radio transmission) in other cases also offline (via docking stations), visually via LCD displays or acoustically (pick-by-voice)
Stationary terminal	Stationary monitors show (online) the pick-up information. Frequently used at central picking points e.g., at goods-to-man picking stations
pick-to-light	Optical display at the shelves show the relevant supply units and the quantity to be retrieved. Frequently used in flow or shelf rack systems.

For this reason, different voucherless methods are used other than the picklist (cf. Table 2.8). These online methods offer the possibility to measure the work progress and thus to set the basis for adapting the order control to the system behavior (system load and capacity). Furthermore, inventory deviations can be recorded immediately (cf. section 2.3.5) and can be integrated into the picking process in due time (on short term).

The voucherless methods offer a better possibility to *acknowledge* orders. When the picked up items are transferred individually [4], each pick-up unit or each item can be checked and acknowledged separately. However, the time consumption has to be taken into consideration which is especially high for the acknowledgement of single pick-up units. A classical problem in pick-to-light systems is that the pickers tend to hit the acknowledgement key *before* the pick what in most cases leads to counting errors.

[4] This is inevitably the case at pick-to-voice and in most mobile terminals; methods like pick-to-light allow one to transfer single or several picking items.

Terminal-based methods offer a more advanced control. Some checks are made with handheld barcode scanners where, for example, the shelf numbers or each pick-up unit have to be scanned. Pick-by-voice systems analogously request check numbers which have to be spoken by the picker.

Despite the different control methods, it is not possible to find a direct connection between error resistance and transfer method. A recent study of picking errors has shown that the picklist picking (voucher) is not inferior to other methods ([53] and Table 2.9).

As the above example has already shown such test methods are first of all meant to support the picker and cannot replace the instruction, participation and motivation of the staff.

Table 2.9. Average error rate in different order-picking systems according to [53]

Technical tools	Average fault rate
Pick-by-voice	0.08%
Voucher	0.35%
Labels	0.37%
Pick-by-light	0.40%
Mobile terminals	0.46%
Mob. terminals + labels	0.94%

Common order-picking characteristics

The basic form of order-picking is shown in Fig. 2.5. The picker moves with his cart along the shelf front (flow, pallet or shelf rack) and grips the units according to the information on his picklist. The cart holds one or several job containers where the articles can be placed separately according to the customer order. Starting at a basic station, B, where he takes up the empty job containers and the picklist(s), the picker starts his tour and finally transfers the filled containers at the interface for shipping. He moves in loops or *meanders* through the aisles (loop strategy). Depending on the picklists single aisles may be skipped or passed only partially (branch-and-pick strategy). According to the terminology defined earlier this is a one-level, order parallel order-picking according to the man-to-goods principle where the picker moves one-dimensionally. The provision is statical decentralized, the transfer of pick-up units is statical decentralized and the transfer of picking units is statical centralized.

In case of large article ranges, there is an alternative as shown in Fig. 2.6 While the principal process is the same the goods are picked in the shelves of a high-bay warehouse. This improves the space utilization and reduces the way

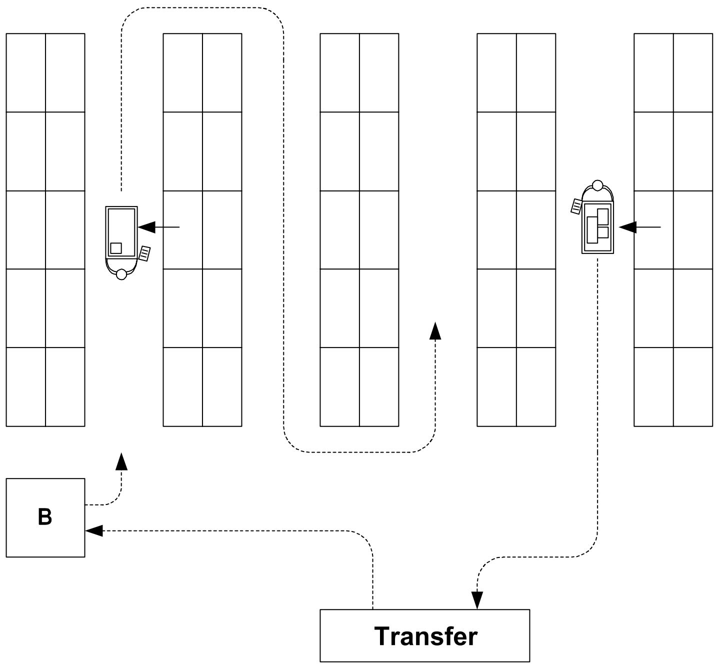


Figure 2.5. One-level order-picking system man-to-goods

times since the movement of the stacker or rack feeder is more dynamical what increases the picking performance. The sequence in which the items have to be picked up from the shelves has a considerable influence on the performance. There are different methods of this so-called trip optimization which are described in Chapter 3. Generally, the picker is guided by a terminal which is installed on the rack.

Fig. 2.7 shows two typical systems which work according to the man-to-goods principle. In the first case (Fig. 2.7, a) the picker stands at the front end of an automatic small parts warehouse (ASW) where the single-item containers are provided for picking by a U-shaped conveying system (roller or chain conveyor). This form is called *picking-U*. In the second example (Fig. 2.7, b), the picker stands at the front end of a horizontal rotary rack where the pick-up shelves are positioned at the front end. Often one picker serves several rotary racks (about three) at the same time. In these cases, the goods are provided centralized dynamically, while the transfer of pick-up and order-picking units is centralized statical. The picker is usually guided by stationary terminals. These systems can be used in one-level as well as in two-step order-picking systems.

The fourth example shows a two-tier order-picking system (Fig. 2.8). The units are picked from flow racks; the supply and retrieval of goods are separated. The picker gets the pick-up information from a pick-to-light system.

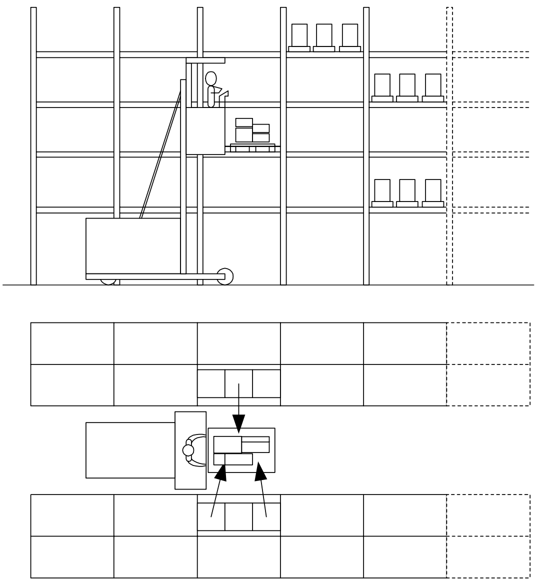


Figure 2.6. Man-to-goods in high-bay warehouse

The retrieved units are directly transferred to a conveyor belt or roller conveyor (pick-to-belt). Thus, the retrieval is decentralized statical, the transfer of pick-up units decentralized dynamical. In the second step the retrieved units are entered into a sorting loop and assigned to the customer orders.

Fig. 2.9 shows a reversion of the common picking process. While in common order-picking the goods are retrieved from a shelf and brought to the transfer station in this case the order containers are arranged on a shelf. Single-item containers are transported from a remote storage area to the pick-up station (decentralized dynamical provision). The units are transferred to the order containers on the shelves which, after completion of the order, are transported to the shipping department by other members of the staff (statical decentralized transfer of pick-up units and statical centralized transfer of order-picking units). This method is called *inverse order-picking* and is of growing importance especially in e-commerce (very large article ranges and many small orders). When an optical device is used to guide the picker this principle is also call *put-to-light*.

Supply control

To guarantee for the smooth and quick picking, the articles have to be available at the pick-up stations. Above all in batch-oriented order-picking systems subsequent picking can delay many orders or lead to incomplete single orders. For this reason, it is of vital importance to control the provided quantities and to initiate the supply in due time.

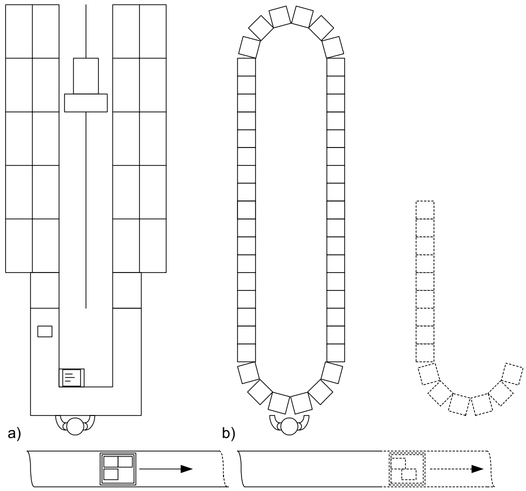


Figure 2.7. One-level order-picking system goods-to-man

2.2.6 Packaging department

In the *packaging department* the provided or picked goods are consolidated according to certain criteria, checked with regard to completeness, packed for transport and finally transported to the shipping department.

In larger warehouse systems customer orders generally consist of sub-quantities coming from different storage areas. Since it is not possible to consolidate such sub-quantities for shipment within a few seconds, the articles are consolidated into transport/shipping units in a first step. This process is often technically supported by order sorting buffers (highly dynamical buffer stores) as e.g., flow racks (cf. section 4.1.3) or rotary container racks (cf. section 4.1.3).

To minimize the shipping costs the consignments have to be optimized. Many companies rely on their skilled and experienced packaging staff to find the right packaging unit (e.g., box size) for each quantity to be packed. In practice, however, it happens frequently that goods have to be repacked — a highly inefficient procedure. For this reason such a function is integrated in most WMS which, based on a volume calculation determine the most suitable packaging. The market offers a variety of computer-aided optimization tools, e.g., for the generation of packaging patterns for palletizing.

This step is terminated during the outgoing goods inspection where it is checked if the customer order is complete and if the quality of the transport/shipping units meets the requirements. This process can be facilitated by consolidating and checking the pickings by means of the recorded order weight as a sum of article and item weights. But this requires exact article master data and mostly homogeneous weights and article ranges.

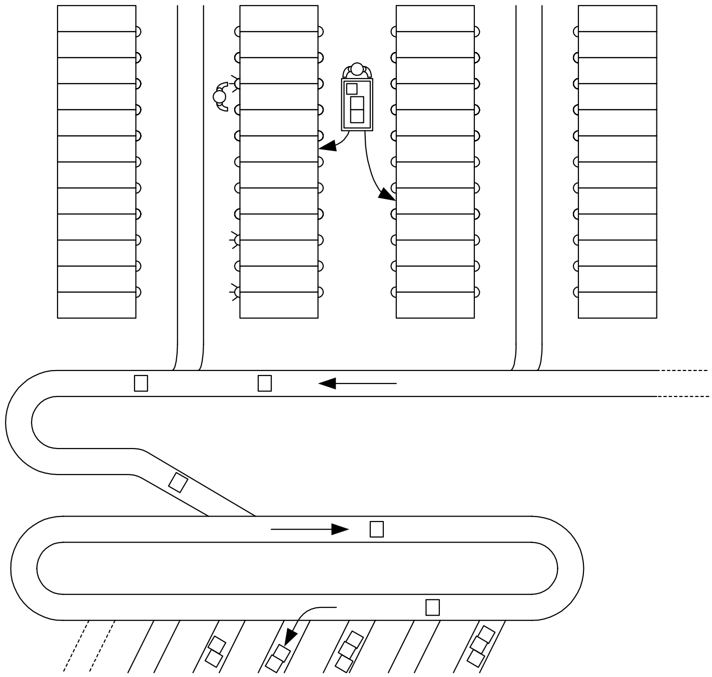


Figure 2.8. Two-tier order-picking system with sorter

Finally, the order status is updated by adjusting the status of the pick-up order and, if necessary, completing the transport and shipping data.

2.2.7 Shipping department

On the face of it, the sole task of the shipping department is to consolidate shipping units according to orders and to load the goods into transport means. Apart from the physical processes in the shipping zone these activities also include a series of organizational and control functions. Unless it is specified in the order the optimal shipping method or the transport means has to be determined. This process is not an easy one due to the different prices of forwarders and CEP service providers. The choice of a cost-optimized shipping requires the exact study of shipping quantities (volumes and weights) as well as of the transport destinations and frequency. Based on these data the optimal transport service provider can be found for each single transport order. At the same time, tours have to be planned where the priority of the delivery, the availability of ordered goods, already agreed delivery frequencies, etc. have to be taken into consideration. These requirements vary considerably according to the storage type and have to be considered in the strategies.

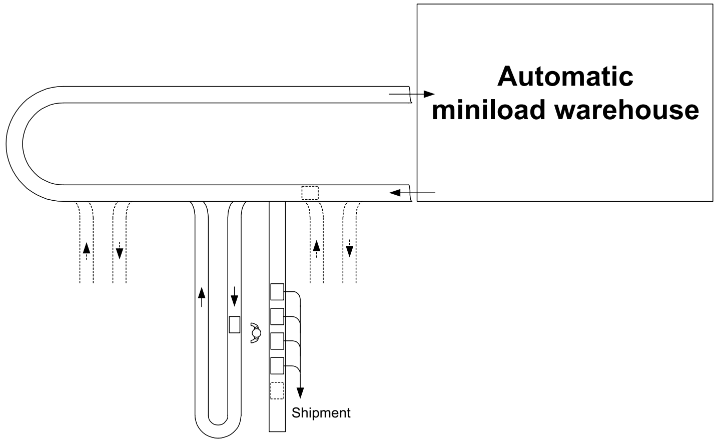


Figure 2.9. Inverse order-picking

While the upstream storage and order-picking area is generally marked by a continuous performance the loading into transport means often takes place within a relatively small time slot, e.g., because of the downstream consolidation of loads in hubs. This leads to receipt and issue peaks which have to be *buffered* in the shipping zones. For this purpose, the shipping units have to be consolidated prior to the loading and provided for loading. When palletized or other larger units are shipped the loads usually are provided in floor storage zones in front of the shipping gate. In addition to this, highly dynamical storage technologies are used to ship smaller units and single items (cf. e.g., section 4.1.3). In practice, the limited space at the shipping gates leads to bottlenecks so that the organization of the shipping zone has to be optimized continuously.

Finally, the transport/shipping documents (freight data, tour-related loading bills) have to be prepared for the shipment of goods. The supply chain can be made transparent by scanning the loaded units. Thus, the finished order is acknowledged and a feedback is given to the order management.

2.3 Warehouse management system

2.3.1 Warehouse management

The management of a warehouse is the main function of a warehouse management system. On the one hand, these systems keep record of the storage capacity, i.e., the specification of the existing storage bins (location management). On the other hand, of the stored units (inventory management). In addition to this, it should also include several control functions to optimize the storage activities.

Table 2.10. Definition of warehouse types

Description	Parameters
Storage bin	Bins, shelves, channels, ...
Access to single locations	Random, stack, LIFO, FIFO ...
Execution	Automatically – manually
Storage operation	Definition of suitable storage facilities (loading capacity, range, rights)

Warehouse type management While the staff in a manually operated warehouse system is able to independently choose the right conveyor and storage facility owing to their knowledge and experience, in an automatic warehouse management system single elements have to be assigned according to their compatibility. Furthermore, the manual operator fulfills certain tasks intuitively while an automatic system is not able to derive the sequence of work steps, like the loading or unloading of the storage channel according to the FIFO principle, from the instructions. Some warehouse functions call for the independent generation of orders, e.g., to define restorage or to optimize the gripping time, and thus for the knowledge of the right storage and retrieval operations.

The basis for such optimizations, thus are the stringent classification of the storage and conveyor technology from the *informational* point of view. For this purpose the warehouse types listed in Table 2.10 should be defined.

Management of storage bins At first, the management of storage bins represents the technical warehouse structure, i.e., the specification of the storage bin based on the storage technique (e.g., shelf racks) including the description of the dimensions, loading capacity and position (e.g., shelf coordinates). Some strategies for the assignment of storage bins (cf. section 2.2.2) require such a precise description of storage locations. More flexible storage methods (e.g., floor storage areas) may only need the specification of the areas and coordinates.

The management of storage bins also includes the management of units stored at a certain location. This includes the entry of goods-specific data such as article specifications (article number or number of the unit load) as well as the registration and update of the quantities stored in each location.

Status data are required for the control of storages and retrievals. When the storage bin is registered at the identification point, on the one hand, its availability has to be sure, and on the other hand, it is has to be ensured that this location is not assigned twice. For this purpose, different states are assigned with regard to the storage bin which is then *blocked* or *reserved* for

Table 2.11. Status information for the control of storage and retrievals (excerpt)

Description	As to location	As to unit load
Available	The storage bin is accessible.	The article is accessible at random.
Reserved	The storage bin is reserved for storage of a certain unit load	The article is reserved for an order which will be executed later. It is ideally reserved with a reference to the order.
Blocked	The storage bin is blocked for future storages (e.g., because of maintenance work).	For some reason (expiry date exceeded, article in quarantine) the article cannot be accessed or is blocked for certain operations (e.g., restorage).

certain articles or orders. In case of a retrieval, it has to be known if a certain unit is available. To make sure that the chosen unit is assigned to the current order the article status has to be linked to the order. The most important states of storage bins and unit loads are listed in Table 2.11.

In addition to the storage and retrieval the *blocking of stocks* or the *setting of blocking indicators* are an elementary management function used for various operations. These are above all

– Blocking for storage and retrieval and
– Blocking for certain warehouse operations (e.g., avoid restorage of fragile goods).

The list of all occupied locations, i.e., the representation of the current warehouse status, is called *bin status report.* The bin status report may also include the type and quantity of articles stored in each bin.

Quantity management (Inventory management) Quantity and inventory management are another logical aspect which focusses on the registration and update of the quantities of each stored article, some times under consideration of the relevant states (cf. Table 2.12). The management of goods according to different criteria (min./max. stocks) should ensure the supply and avoid excess quantities. When fixed limits are exceeded or not achieved messages or actions (orders, restorage, etc.) have to be generated.

This function, however, requires a careful control of the stored goods including the allowed storage time and blocking of the article when a certain (expiration) date has been reached. Under certain conditions the respective goods have to be taken out of store to protect other goods.

The main difference to an *enterprise resource planning system* (ERP), which has similar functions, is the warehouse management, while an ERP

Table 2.12. Inventory categories

Description	Meaning
Physical stocks	Units available in the storage system
Available stocks	Stocks under consideration of blocked or reserved quantities = inventory of available units
Reserved stocks	Inventory marked with a blocking indicator
Shortages	Open (outstanding) incoming deliveries for whichan order has already been submitted.

system mainly focusses on customers and sales. For this reason, a warehouse management system generally does not contain customer data or prices. Nevertheless, a functioning overall system needs the continuous exchange between WMS and ERP.

System control The control of the warehouse condition (temperature and humidity) and security (access control) is a special function which is of vital importance only in a few cases.

Groupings

In a warehouse some regularly occurring tasks refer to a group of unit loads, articles, shelves, etc. Therefore, a warehouse management system allows for flexible groupings in order to avoid the labour-extensive handling of single elements. Such tasks occur, for example, during:

– the maintenance, repair or failure of single aisles in an automatic shelf system,
– the goods control (e.g., special measurements for hazardous or high-quality goods),
– the storage or blocked goods (e.g., quarantine storage),
– the choice of larger quantities or
– stocktaking.

Ideally, such groupings should be highly flexible, but in any case it should be possible to group

– storage locations (according to aisle, zone, etc.)
– article groups (according to type, article number, etc.)
– batches

As a consequence, it should also be possible to apply the management functions (blocking, reservation, storage/retrieval) to the complete group.

2.3.2 Reorganization

An active warehouse and distribution system should be checked regularly with regard to its efficiency and suitable steps should be taken for its optimization. This is called *warehouse reorganization.* This action can be initiated by:

– changed retrievals of certain articles, e.g., lower throughput or change of of the typical pick-up unit
– beginning or terminated campaigns
– changed product range
– growing number of partial storage units

As a result, storage areas are occupied faultily (e.g., wrong zone), the average transport routes increase and the utilization of space is reduced. For this reason, the corresponding parameter should be analyzed and controlled continuously. The suitable system control method is described in section 2.3.4.

The resulting optimization of the system is achieved by some measures which logically should be supported by the warehouse management system:

– rebooking, i.e., reassignment of articles to suitable retrieval categories and storage zones
– restorage of already existing units in times with low retrieval rates (e.g., to reestablish an ABC zoning)
– densification of partial units or badly utilized mixed pallets (retrieval of relevant units, repackaging of units and restorage according to specifications)

2.3.3 Conveyor management and control systems

While automatic conveyors are controlled and supervised by a corresponding control (material flow controller, MFC), manual unsteady conveyors (stackers, etc.) are controlled by different systems from the manual management to the fully automatic system control.

Computer-aided control systems for the in-house transport are used in warehouse management for different reasons:

– To optimize the system performance (reduction of empty trips, larger handling volumes, higher system load)
– Flexibility of the system in case of short-term changes (quick reaction to transport requirements)
– Control of the system status (runtime of vehicles, operational costs per vehicle, etc.)

Systems for the deployment and control of vehicles are called *stacker control systems* or *transport control systems* [90]. These systems consist of a computer-aided control station or master computer, a wireless transfer medium (radio or infrared) and mobile terminals on the vehicles. Incoming transport orders or requests are handled, completed by relevant data (e.g., completion of article number or description, storage location (source), destination (sink)) and transmitted to the driver by certain procedures and strategies.

In all control systems first of all it has to be decided which conveyor has to be assigned to a certain kind of order because of its loading capacity, lifting height or other specifications. For this purpose, a classification of the existing conveyors has to be available in addition to the management of warehouse types (cf. section 2.10). Furthermore, the restrictions of other facilities like transfer points have to be recorded. There are two principle methods for the vehicle deployment:

Dispatching In dispatching a suitable vehicle is assigned to the current order according to different criteria and strategies. This may be, for example,
- the next free conveyor
- the nearest conveyor
- the conveyor with the shortest connection trip

Since the dispatching handles the currently incoming order the system reacts quickly and flexibly and the system can, thus, be used in dynamical environments.

Scheduling In scheduling, on the other hand, several orders and/or conveyors are assigned to an “ideal” schedule. This method thus aims at an optimal order sequence and system performance. The main condition for a scheduling is the collection of queued orders in an order pool from where they can be assigned optimally. In contrast to a manual scheduling, which can seldom be used in a warehouse, the scheduling is repeated in relatively short intervals.

In practice, mostly a mixed form of both methods is used where the most important requirements can be considered at any time. In addition to these extensive dispatching control systems there are simple conveyor management systems which focus on the recording of the system status and where no orders are assigned. This includes the recording of operating times or repair costs per vehicle, the supervision of the battery status and control of maintenance intervals for each vehicle. In systems where the vehicles are used by a large number of people it may be reasonable to also record and document the driver identification.

2.3.4 Data collection, processing and visualization

As already described in section 2.4 a warehouse and distribution system includes a variety of different data and key values for different purposes:

- Activity recording
  - Customer service
  - Documentation of shortages from stocktaking
  - Recording of picking errors, deviations at dispatch control
  - Efficiency of the staff (e.g., picks or orders per picker; storages per stacker driver; waiting times per vehicle)
- Overview
  - Bin status report (sorting according to storage bin, free/occupied, ...)
  - Stocks
  - Warehouse statistics (transshipping frequency, error times, filling level, ...)
  - Utilization of space
- Operating means statistics
  - Runtimes
  - Idle times
  - Maintenance and repair costs per unit
- and much more.

While the personal activity recording is not without legal problems and generally requires the admission of the works council it sometimes is the indispensable basis for the activity-based accounting especially with regard to

- the recording of piece rates
- the recording of contracted logistic services

The recording of significant status information is the elementary prerequisite for the control and optimization of the distribution system. Based on these data manpower requirements (e.g. in order-picking) or other resource requirements have to be determined and scheduled in the system control. The loading levels of conveyors show if parallel arranged systems are balanced. Waiting times at certain points may indicate bottlenecks and initiate a check of the operating strategies, system performances or personnel deployment.

The recording method is decisive for the use and successful evaluation of the data. There are two principle methods to record significant data and key values:

Online recording The database which is necessary to generate the required data is recorded in the process and automatically converted into the desired key value. Thus, the key values are directly available. The recording and evaluation focusses on the predefined problems. Other or similar data and key values cannot always be evaluated afterwards. The recording time is almost fixed and cannot be reset, for example to block out certain events or periods, because the results just are aggregated key values. Thus, this method represents a statical system.

Time series recording At first, only a *log file* is made out containing the events and the time of occurrence (time series). Here, it has to be ensured that the input data are determined purposefully. The required key values are then extracted from the database via a corresponding request. This method clearly offers much more possibilities and a better basis for the system planning and optimization. However, the data volume may increase continuously. It is advisable to book the following data:

- Documentation of requested and completed orders
  - Requests:
    - Request ID/demand ID
    - Source/sink
    - Date/time
    - Status (express-normal)
  - Fulfillment:
    - Person or operating means ID
    - Date/time
    - Termination
- Operating protocol
  - Start/end of operation
  - Error message and times
  - Warehouse movements
- Single information
  - article-related data (article no. stocks,...)
  - order-related data (order no., order item, dates...)
  - unit data (storage bin no., free/occupied, quantity ...)

In addition to this, delivery notes and goods issue protocols have to be recorded and documented for legal reasons.

2.3.5 Stocktaking

In most countries stocktaking is a legal obligation which has to be carried out by each businessman for each business year. For fixed assets and supplies, i.e., stocks, a *physical stocktaking* has to be made. This has the aim to check the inventories (booked stocks) and the reliability of the inventory management (warehouse accounting). All objects (storage units) have to be identified and classified by counting or measuring. The data have to be recorded in a protocol showing the following information:

- Voucher no. (check of completeness)
- Storage location and position
- Description of the object
- Recorded quantity and quantity unit
- Unit price and total value
- Optionally information about value-related influences (age, storage time)

– Date of recording and signature (or recording person)

This procedure is immensely labour-intensive and nearly impossible in very large warehouses. Furthermore, certain automatic warehouse techniques do not allow for direct personal checks and for the retrieval of all units for process-related or economical reasons. For this reason, different stocktaking methods have been developed which have to be cleared, however, with the respective auditors and financial authorities.

Annual stocktaking The classical form of stocktaking requires the physical check of all stocks at the accounting date. Since there is no business at this day there are no changes of the "assets". Such a method is only suitable for smaller systems. It is not necessary to finish the stocktaking within one day but as close to the accounting day as possible, i.e., 10 days prior or after the date, as far as inventory changes are recorded and considered in the stocktaking.

In a *forward or subsequent stocktaking* stocks can be taken at one day, three months prior to or two months after the accounting day and recalculated by a *special stock* . This requires the use of an updating or backward projection according to the *Generally Accepted Accounting Principles GAAP.*

Permanent stocktaking The permanent stocktaking is a suitable method in times of little business or low stocks. Stocks can be physically taken throughout the year if all inventories are booked continuously and all receipts and issues are booked separately mentioning the day, type and quantity. At the accounting day a (quantitative) inventory update is made — a so-called book inventory.

The counting may be related to the articles or the storage bins. For this purpose, all movements of the respective article or storage bin must be stopped. For an article-related stocktaking all free storage bins have to be checked separately. For this method the EDP-aided warehouse management system has to be able to assign article or bin-related counters.

Since in automatic warehouses the stocktaking cannot be performed at the storage bin for technical reasons the most suitable methods are the warehouse stocktaking and the zero-crossing stocktaking. During a warehouse stocktaking the items are counted at another location (usually at the identification point prior to the storage) where they are marked with the counter. In a zero-crossing stocktaking all storage bins are recorded at a zero-crossing (the stocks in the bins are completely removed). Differences are entered directly and the counter is updated[5].

Inventory sampling If an EDP-based warehouse management system is used and the Generally Accepted Accounting Principles are met an in-

[5] If shortages are recognized the warehouse management system has to initiate a replenishment independent of the requirement (inventory) or has to consider other retrieval points to fulfill the customer order correctly.

ventory sampling may also be used which requires considerably less expenditure. Samples are taken physically which are evaluated by means of proven mathematical statistical procedures, i.e., above all:

Sequential test The number of samples is unknown at the beginning and results from the repeated check of the test criterion. The test is repeated until the acceptance criterion is fulfilled (fall short of minimum error rate) or rejected (exceeding maximum error rate). To limit the testing period a stop criterion may be defined.

Estimation The basic total is estimated based on the frequency ratio of a sample. In case of *layered* estimations the basic total is splitted from which a sample each has to be taken. *Bound* estimations also use an auxiliary parameter, e.g., the booked inventory value, for the projection.

As already mentioned above, the chosen method has to be accepted by the auditor or financial authorities. The following parameters influence the suitability of a method:

- Use or existence of an EDP-aided inventory management system
- Accessibility of the shelves (freely accessible warehouse or closed area of an automatic warehouse)
- Value of the goods (The more valuable the goods the more exact the stocktaking. Sampling is unsuitable for "high-value goods")

Warehouse management systems should have at least the following stocktaking functions:

- Counting date for storage units and shelves
- Blocking of article groups or shelf areas for stocktaking purposes
- Permanent updating of counters considering the recording person, date and time
- Zero-crossing stocktaking

The rigid stocktaking requirements do not come up to the possibilities of modern warehouse management systems. A continuous counter booking and zero-crossing comparison allows for an extremely detailed stocktaking. And it should not be ignored that errors during the stocktaking process can hardly be avoided.

On the other hand, an exact inventory management offering safe data is of vital importance for the readiness to operate. The combination of a regular data update and a correct data management set the basis for a high readiness to deliver and short response times. Furthermore, negligence or thefts are detected only by a reliable stocktaking.

2.4 Basic data and key performance indicators of warehouse systems

The planning and design of warehouse and distribution systems is highly complex so that it is nearly impossible to describe all relevant system-related parameters. The major part of such key values has to be defined for a given problem. In the following the most elementary key values which are used in quite a number of systems are described. They are classified into *basic data* and *key values*

Basic data are also called absolute numbers and directly result from measurements, counts, summations or differences of certain units or they are recorded as master data. At the same time, they also represent the requirements and basic information to be provided by a system.

Key values should provide significant and densified information to evaluate and compare the efficiency of processes and systems. Here, absolute as well as relative numbers, i.e., proportionate values or data are used.

2.4.1 Basic data

Master data

Master data are statical data which are not changed over a longer period of time. The master data contain all important information about the basic characteristics of an article, loading aid, etc. The most important master data in a warehouse are the article master data because all main warehouse functions and control mechanisms are based thereon.

The article master contains a description of all articles independent of their current stock. The total number of articles represents the assortment although discontinued or *dead* articles result in differences to the actual stock. Table 2.13 gives an example of the most important elements of the article master data.

Inventory data

This data group informs about the quantities of articles stored or provided over a longer period of time. The up-to-dateness and accuracy of this kind of data collection is of special importance to ensure the readiness to deliver and to dimension the warehouse system. Since these data change continuously they are also called dynamical data.

Movement data

The second group of dynamical data are the movement data which represent all main physical warehouse processes. These are basic processes like goods receipt and issue and warehouse operations as well as order-picking processes and order execution.

Table 2.13. Elementary basic data

Article master data	Inventory data	Movement data	Other System data
– Article number – Description – Article weight – Article length – Article width – Article height – Quantity unit – Type unit load – Loading factor (packaging quantity/unit load) – Gripping unit (packaging quantity/retrieval unit) – Blocking indicator – ABC-classification – Batch number – Weight/retrieval unit – Weight/unit load – Client – Best before date – Remaining runtime – Sorter capability	– No. articles – Total stock – Average stock – Minimum stock/art. – No.UL/art. – Available stock – Shortages	– Goods receipts/ day – Goods issues/ day – Storages/d – Retrievals/d – Quantity transship./a – Restorages/d – Orders/d – Orders per article – Positions/order – Positions/d – Grips/ Pos. – Incoming orders/h – Order lead time – Material lead time – No. of orders/ order type – Double cycles/d – Complete units/d	– Order types – Unit load master data – Packaging master data – Storage capacity – Space restrictions – Room restrictions – Utilization space/volume – No. UL/art. – No. staff/dept. – Sick days – Operating costs (manpower, energy, maintenance) – Investment costs (replacement) – Value turnover/a – Productivity

Other system data
Other elementary system data are among others

– Structural space and room data
– Structural manpower data,
– Cost data
– Unit load and packaging master data, etc.

2.4.2 Logistic key performance indicators

As already described in the preceding section a variety of different information arise in warehouse and distribution systems in the form of data. Because of this abundance it is difficult to evaluate and optimize the system. Furthermore, some data may be misleading when taken out of their context. The *No.*

orders/d, for example, reveal very little when it is not seen in combination with *No. items/order*.

In a narrower sense key performance indicators are densified parameters, i.e., values calculated from data and other key performance indicators. On the other hand, all kinds of parameters with the following characteristics are summarized as key performance indicators:

> Logistic key performance indicators are numbers by which the quantitative aspects of logistics can be described in a concentrated form [66].

Accordingly, basic data also are key performance indicators. The specialization of logistic key performance indicators takes into consideration that key performance indicators are used in all areas of technology, economics, etc. Key performance indicators are used to give a quick overview over optimal costs and performances [66] and to evaluate the different variants.

Above all *efficiency/productivity values* (output/input) and *intensity values* (input/output) are built to derive relative key performance indicators. The generation of specific key performance indicators depends on whether the problem is of an operative or a strategic nature. Operative key performance indicators are first of all used to control efficient logistic processes while strategic key performance indicators are used to develop and design efficient goods flows [28]. Key performance indicators are often based on averaged and approximated values and provide no precise information but just a quick overview. Typical key performance indicators are shown in Table 2.14.

The key performance indicators "quantity turnover" and "inventory coverage" show that single key performance indicators may be defined differently what leads to considerable evaluation discrepancies. Therefore, it should be checked carefully if a key performance indicator relates to the value, quantity or performance.

Since single key performance indicators can represent only partial aspects and the variety of possible key performance indicators and their combinations make it difficult to use them targeted and systematically they are combined in *key performance indicator systems* [66]. Thus, the auxiliary key performance indicators which contain the necessary information for a management task are combined in a hierarchical structure of systematically linked single key performance indicators. According to Reichmann, the top key performance indicators of such a system are the transshipment frequency, the overall logistic costs and the readiness to deliver. The main problem, however, is to develop such a system for a special application.

While single key performance indicators can easily be used to analyze deviations and to reach certain goals in warehouse management extensive key performance indicator systems are a main tool of logistic controlling and will not be described here any further.

Table 2.14. Examples of logistic key performance indicators

Key performance indicator	Definition	Objectives/ problems	Influencial variables
Readiness to deliver	$\frac{\text{no. of requests delivered in due time [pc]}}{\text{no. of requests [pc]}}$	Delivery service, customer satisfaction	Order handling, throughput, capacity
Warehouse fill degree	$\frac{\text{number of occupied locations [pc]}}{\text{storage capacity [pc]}}$	Utilization of storage capacity	Inventory management
Transshipment level quantitative	$\frac{\text{retrievals [pc/a]}}{\text{storage capacity [Stck]}}$	Warehouse dynamics	Size of shipping units (because of customer orders)
value-based	$\frac{\text{total turnover [\$]}}{\text{Ø stored value [\$]}}$	Inventory cost	Ordering system
Cost / storage location	$\frac{\text{total cost [\$]}}{\text{storage capacity [pc]}}$	Choice of warehouse technology	Warehouse technology
Inventory coverage quantitative	$\frac{\text{current stocks [pc]}}{\text{wareh. turnover [pc/a]}}$	Inventory costs, wareh. dynamics, service level	Ordering system
value based	$\frac{\text{stocks [\$]}}{\text{wareh. turnover [\$/a]}}$	Inventory costs, service level	Ordering system
Picking way/item	$\frac{\text{average picking route [m]}}{\text{Ø line items [pc]}}$	Choice of picking technique	Process control, picking principle, information management
Picking density	$\frac{\text{no. of line items [pc]}}{\text{gripping space}[\text{m}^2]}$	avg. waytimes, picking performance	Supply management, rack technology

2.5 Special procedures and methods

2.5.1 Cross docking

In cross docking goods receipts and dispatches are coordinated to such an extent that incoming goods are directly transferred to the shipping department without being put away. Thus, they are not entered into the warehouse system which is a pure transshipping system with the aim to:

– reduce stocks at certain points along the supply chain
– increase the efficiency by avoiding process steps
– reduce the throughput times of articles in the system
– improve the services by more frequent deliveries
– sort articles effectively according to destinations, e.g., for CEP

This method should not only minimize the stocks in the distribution system or at the *cross docking point* but also at the POS. This requires frequent deliveries what improves the services. There are two principle cross docking methods:

1. Cross docking with broken pallets
2. Cross docking as flow system

Cross docking with broken pallets
The incoming units are quasi single-item units and have to be distributed or picked according to the orders of the single subsidiaries. Palletized units are typically transshipped in roller containers, a principle which is also called *two-level* or *container cross docking*. A main feature of this method is an order-picking process.

Cross docking as flow system
Here, the supplier pre-sorts the incoming units according to the orders of the subsidiaries in such a way that the single units do not have to be broken and distributed but only have to be consolidated, i.e., combined with other single-order units. This principle is also called *one-level cross docking*. If only complete transport units (e.g., pallets) are handled this method is called *pallet cross docking*. If the pallets hold pre-sorted containers which have to be assigned to the single subsidiaries or tours this principle is called *pre-sorted store order*. In this case the goods are only transshipped without counting or direct picking processes.

Prerequisites and fields of application Many requirements have to be met before the cross docking principle can be used. One efficient way is to directly transfer incoming goods. However, this requires that the desired quantity is available on short notice so that the risk of shortages is very high due to lacking stocks. It is practically impossible to restore articles in case of an order cancellation. If the subsidiaries have similar order structures the tours are scheduled almost simultaneously. This may lead to bottlenecks at the goods issue and thus to several deliveries of single subsidiaries a day.

In practice this method is only suitable for systems with a relatively constant demand, similar quantities and articles, short replenishment times and

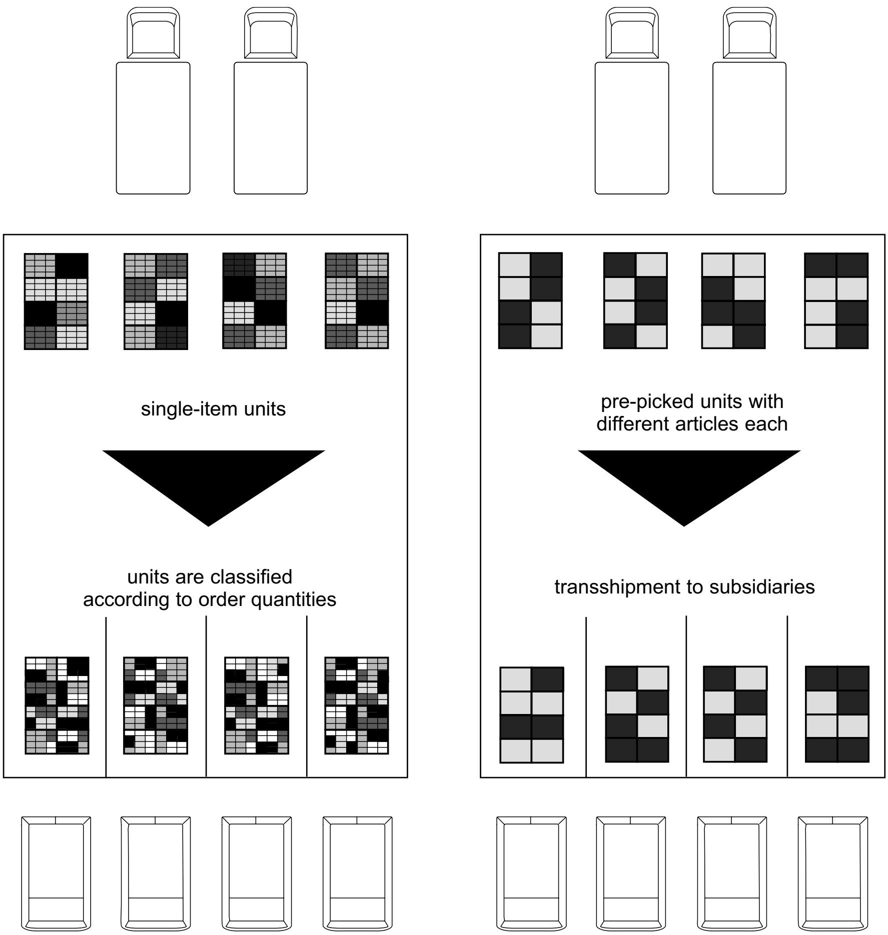


Figure 2.10. Cross docking principles

short distances to the subsidiaries, for example for fresh goods. Furthermore, these principles are used for certain groups of goods within existing goods distribution centers where corresponding processes have to be integrated into the warehouse management system. [11] gives recommendations for the realization of communication interfaces.

2.5.2 Outsourcing of the physical distribution and warehousing processes

The outsourcing of logistic services was already described earlier in this book (cf. p. 15). The warehouse management is influenced, among others, by connecting external systems to the own enterprise resource planning (cf. Chapter 7).

To be able to adjust the processes in multi-client outsourcing warehouses to the special requirements of the customers the processes, procedures and strategies have to consider not only the goods and customers but also the client. This feature has to be integrated into the WMS which is then called *multi-client enabled*.

Since provided services are often settled on the basis of performed activities single client-related services (e.g., stacker trips, picking positions) have to be recorded separately.

2.5.3 Application Service Providing

Similar to the outsourcing of physical warehouse processes to an external service provider in Application Service Providing (ASP) IT services and computer-aided control functions are outsourced to a service provider who controls the processes from a central computer center where he also carries out the necessary evaluations, processing, etc. Such services are successfully used in wage and financial accounting.

An advantage are not only direct cost savings, e.g., by a reduced EDP staff on the site, but above all the data security. A central service provider can afford other better possibilities than many companies at decentralized locations.

In logistics such solutions are used in the form of shop systems in e-commerce and ERP but also on WMS superordinate levels. An ASP solution where time-critical processes in a large distribution system are controlled from a central computer center is still unknown because the transmission security cannot be guaranteed and the response times may be too long. Breakdowns and failures would inevitably lead to process interruptions.

With regard to current developments in the information and communication technology such solutions are quite conceivable above all for smaller applications. The basic functions remain unchanged so that the special procedures will not be described here.

3. Fundamentals of an Operational Optimization

The *optimal* operation of a warehouse or distribution system has been achieved when every customer order is ready for dispatch completely and in due time and when all necessary processes are carried out within the shortest possible time and with minimum resources also under changing conditions. To meet these requirements the processes have to be planned and coordinated in advance.

The order planning and dispatching in a warehouse management system assigns all goods movements in the warehouse to the respective resources and determines the time and order of the order execution as to avoid delays and to achieve a continuous system load without bottlenecks. This also helps to minimize unproductive auxiliary process times.

This chapter describes the methods of a computer-aided order dispatching and the basic *algorithms* and *heuristics* of a typical warehousing problem.

3.1 Optimization in short

The achievement of economical, technical and organizational goals requires optimized operational processes in a warehouse or distribution system.

Each work step which has to be performed in a warehouse requires one or several resources, i.e., personnel and operating means which in turn consume money and time. The aim is to provide the maximum transshipment performance consisting of all single work steps with the lowest possible amount of resources.

Furthermore, the relationship between supplier and customer results in additional frame conditions, which have to be considered when operational processes are optimized: agreed time slots for deliveries and shipments of ordered goods define the period of time for the execution of certain activities. The *ability to provide information* to the customer about goods and delivery or collection dates is only possible with a preceding scheduling.

3.1.1 Background

Optimal acting means to choose “that decision out of a number of possible decisions which best meets a specified target” [61]. Such situations occur in all areas of our life. In logistics decisions closely linked to make the optimal choice. A typical example is the planning during the design of a material flow. Different system solutions are developed which have to be limited to one variant during the selection process. This choice is based on the functions of the material flow system to be designed.

During the operation of a material flow system a very large number of decisions has to be made continuously, e.g., concerning the order dispatching where the optimal order sequence has to be determined.

Optimization is *the* classical task related to the design and operation of all technical systems. It comprises all aspects of the system design with the general aim to provide a desired function at the lowest possible overall costs. Consequently, there is a large variety of rather diverse methods which all together are called the *optimization theory*.

At first, this chapter gives a short overview over the optimization theory followed by the classification of warehouse problems.

Principally, optimization is the same as the targeted search for the optimum result within a given scope of action or within a number of given alternative actions. The aim is to find that one alternative offering the optimal result in one of the following scenarios [41]:

– To achieve the best possible result at a given expenditure (maximization problem).
– To achieve a defined result at the lowest possible expenditure (minimization problem).
– To achieve the optimal result at the lowest possible expenditure (so-called dual problem).

In order to determine the optimal solution the alternatives have to be *quantifiable*, i.e., comparable. An important element of an optimization is the definition of target criteria as well as of a *target function*, as basis for the determination of the *achievement degree* or *optimum* of individual alternatives. Thus, the basic task of all optimization activities is to

> Determine the aim (W) under consideration of the given situation (G) so that a target function $F(W, G)$ is either minimized or maximized [61].

Many scientific disciplines have contributed to the development of user-specific mathematical optimization methods. These are, among others, the *cycle theory* and the *decision theory* [65] which, however, are mostly used to solve *economical* problems.

Two scientific sectors with a different history directly deal with the planning and operation:

Operations research (OR) aims at developing methods and procedures (algorithms) for the determination of optimal solutions. Typical fields of application are the preparation of time schedules and the *sequencing* of work steps as well as the solution of standard problems like the minimization of trips in a warehouse by an adequate arrangement of storage areas. Another important criterion of OR is the expenditure to be spent on a solution. As described later on in this chapter, there are optimization problems which cannot be solved optimally or not at all by computers within a finite period of time.

Artificial intelligence (AI) goes beyond the methods of OR since it considers the aspect of an autonomous decision making for machines (computers) and tries to transform patterns of the human intelligence to the computer. An example is the modelling of certain optimization problems by means of artificial neuronal networks. Here, the main emphasis is given less to the determination of an exact optimal solution but to the general solution within an acceptable period of time even if this would mean a loss of quality.

Because of partly identical applications and methods both disciplines can hardly be differentiated; both are computer-aided methods and thus part of the *computer sciences.*

3.1.2 Classification of the operational optimization

During all phases of a warehouse and distribution system, i.e., from the planning up to the operation there arise problems with regard to optimization. The general goals of an optimization are:

- Consistent load of the operating means and staff
- Avoidance of unnecessary idle times
- Reduction of empty conveyor trips
- Minimization of order lead times
- Observance of given order time slots

A first approach is to define a time horizon for the following problems *Long-term* planning tasks deals with the design of a warehouse system in line with the *strategic* problems which have to be solved, e.g.:

- Dimensioning of storage areas and choice of the warehouse technology.
- Preparation of a warehouse layout (arrangement of the storage areas and access ways).
- Development of fundamental operating strategies and proof of the system functionality, e.g., by simulation.

In a warehouse layout the storage areas and ways have to be arranged in such a way that the forecast goods flows can be handled by the chosen system technology with minimum transport expenditure. Although the planning of the facility is not part of the operational optimization process a detailed knowledge of the *operating strategies* to be used may be required to operate the different working areas of a warehouse system. To prove the overall functions, the operating strategies which have to be implemented into a WMS could be represented in a simulation model and thus be optimized at an early stage.

Medium-term planning processes, which are due, for example, monthly or weekly, are based on the turnover of a warehouse system during this time and have to adapt the available resources to seasonal fluctuations. Typical planning tasks are:

– Planning the assortment, adapting the article range to be kept in stock in order to react to seasonal or market-related influences.
– Manpower requirement planning, preparation of shift models or work schedules.
– Reorganizing stocks, e.g., stock transfers following an ABC analysis[1].

The reorganization of stocks by means of article transfers with frequent access to the front storage areas also is a method to improve the access on an article and thus to minimize order lead times. If sufficient capacities are available for the execution of orders this method can also be applied in shorter intervals. But the resulting savings have to be seen with regard to the additional expenditures for stock transfers.

Organizational optimization in its real sense means the *operative planning*, i.e., the management of daily work processes. It is based on the current order and delivery volume detailed to the day and represents the operational activities over a maximum period of just a few days or hours. It has to fulfill the following tasks:

– Resource assignment, scheduling and sequencing of picking and transport orders
– Planning of transport routes within the warehouse
– Comprehensive planning of *batches* , i.e., preplanning all work processes in the warehouse

The period of time which is available for the management varies according to the tasks and is furthermore influenced by short-term changes like express orders, unplanned goods receipts (deliveries unscheduled by the supplier, wrong quantities) or operational breakdowns.

Depending on the available data of the orders to be managed and the available period of time, the optimization times are still classified into *offline-* and *online problems.*

[1] The generalized ABC analysis is a so-called *cluster analysis* where access frequency on the available articles is not analyzed for three (A-B-C) but for any number of articles. In the most extreme case the quantity to be studied is just one article.

Offline problems are marked by the fact that all order information are available before the dispatching, e.g., the calculation of an optimal (*order schedule* cf. section 3.2) The offline scheduling is the classical form of sequencing in OR and is described in detail in the relevant literature.

Online problems occur when orders are assigned to the operating means without a scheduling or when new orders are assigned to a running work processes with a fixed procedure. Online problems are also called *dispatching*.

Online problems also arise when new orders have to be added at short notice to an already planned schedule because of unannounced deliveries or express orders.

The optimization theory comprises quite a number of other problems and optimization models which differ only slightly with regard to their logistic tasks and will not be discussed any further in this book. .

3.1.3 Terms and elements of dispatching

According to [71] possible dispatching methods are divided in

- **Order-based** dispatching, i.e., a still unscheduled order is chosen which is then completely scheduled, i.e., suitable resources and intervals are chosen for every work step. This is repeated until all orders are scheduled.
- **Resource-based** dispatching, i.e., one resource is chosen for the scheduling of the best suitable work step. This is also repeated until all orders (all order steps) are scheduled.
- **Operation-based** dispatching, i.e., an operation (work step) and the suitable resource and interval are chosen until all orders are scheduled.

The remarks on the optimization of order dispatches have to be considered as operational dispatching. For an exact classification and representation of the dispatching processes in the form of a batch generation please refer to section 3.2.4.

Customer orders are orders managed by a subordinate ERP system and contain information about which article has to be provided at a fixed date at the goods issue in which quantity for which customer. The basic elements of a dispatching are *orders* which have to be executed by the warehouse *resources*. Unlike customer orders, which initiate a warehouse operation, the term order describes the movement of just one storage unit (loading means, package, article) from a *source* to a *sink* and thus an elementary logistic operation. Order elements which have to be considered during the dispatching are, for example:

– Clear labelling of the units to be moved (LHM-ID, article number)
– Source and sink
– Executing resources
– Order reference (e.g., sub-order of a picking order)
– Time slot (earliest starting date, latest termination date)
– Priority
– Order status (still unprocessed, in process, finished)

Depending on the type of order sources and sinks may be storage locations, resources or loading facilities. Together with the choice of the executive resources and information about the storage units to be moved they constitute the main order data.

Resources are the executive units within a warehouse system, i.e., the manpower and the conveyors and storage facilities. To be able to assign the orders to be executed within a certain period of time to the warehouse resources information about these resources have also to be available, i.e.:

– Performance (conveying speed, gripping time, etc.)
– Capacity
– Current work load

In addition to this, order dispatching requires information about the topology of a warehouse, on the one hand, to calculate the order execution time based on the covered distances and, on the other hand, to determine the shortest route in case of several possible routes between source and sink. The presentation of routing or cycle times in the form of *adjacent matrixes* is a necessary element of the WMS data model.

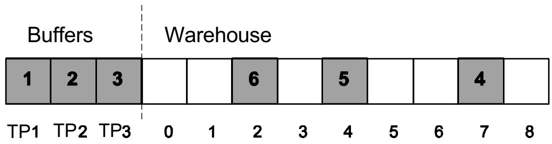
3.2 Optimization processes in a warehouse

At first, the methods of an operational optimization and the related problems are described by use of some examples.

3.2.1 Transport optimization

A WMS controls the transport within a warehouse, e.g., pallet transports between the different storage areas with stackers, transports on conveyors or the storage and retrieval in automatic high-bay warehouses, which will be described in more detail below.

In a high-bay warehouse, all transports are managed by the WMS as single transport orders for one loading facility (cf. Chapter 5) stating the source and sink coordinates. Then they are transferred to the subordinate control of the rack feeder in the warehouse aisles. Thus, the WMS controls the order of cycles to be performed by a rack feeder in an aisle. In each aisle



Order	**Source**	**Sink**
1	TP1	SL1
2	TP2	SL3
3	TP3	SL6
4	SL7	TP1/TP2/TP3
5	SL4	TP1/TP2/TP3
6	SL2	TP1/TP2/TP3

Figure 3.1. Example of a double cycle in a high-bay pallet warehouse

the loading facilities to be stored as well as those retrieved by the rack feeder are buffered in storage bins in the warehouse prestorage zone. Thanks to this buffers the transport movements in the HBW are decoupled from the remaining warehouse material flow.

A basic optimization method in an aisle is to combine simultaneous storages and retrievals into *multiple cycles*, where one empty rack feeder makes the connecting trips between a storage and a retrieval point to minimize the routes and thus to maximize the throughput.

Creation of double cycles

In the present case, the load-carrying equipment of the rack feeder can take up one pallet. Since the prestorage zone as transfer point has to be approached at the beginning and the end of a multiple cycle the optimization is limited to combining two transports, a storage and then the retrieval, into a *double cycle*. If there is a sufficient number of transport orders the rack feeder starts the double cycle by taking up a pallet in the prestorage zone and stores it at the designated location. Then, it approaches a pallet to be retrieved and transfers it to its location in the prestorage zone.

If the order list of the WMS holds more than one storage and retrieval job for an aisle which can be executed in complete double cycles without sequencing restrictions the WMS determines the best possible combination of double cycles before charging the rack feeder.

Table 3.1. Determination of the optimal sequence of the three double cycles

No.	**Double cycle (Order sequence)**	**Distance (Width of storage location)**
1	1 - 4 *(TP1 → SL1 → SL7 → TP1)*	20
2	1 - 5 *(TP1 → SL1 → SL4 → TP1)*	14
3	1 - 6 *(TP1 → SL1 → SL2 → TP1)*	10
4	2 - 4 *(TP2 → SL3 → SL7 → TP2)*	18
5	2 - 5 *(TP2 → SL3 → SL4 → TP2)*	12
6	2 - 6 *(TP2 → SL3 → SL2 → TP2)*	10
7	3 - 4 *(TP3 → SL6 → SL7 → TP3)*	16
8	3 - 5 *(TP3 → SL6 → SL4 → TP3)*	14
9	3 - 6 *(TP3 → SL6 → SL2 → TP3)*	14

The order volume assumed in Fig. 3.1 consists of three storage and retrieval jobs and thus of three double cycles[2], which have to be executed by the rack feeder in a consecutive order. At the beginning of the sequence the transfer points TP1 to TP3 are occupied by the pallets for orders 1 to 3. The pallets which are retrieved in the course of a storage process are then stored at the released storage locations (TP1, TP2 or TP3).

The first of the three storage jobs 1 to 3 may be followed by one of the three retrieval jobs 4 to 6. The following storage job (one out of two) may then be assigned to one of the remaining two retrieval jobs. The last double cycle consists of the remaining two orders. Since each of the storage jobs 1 to 3 can be the first in this sequence the number of possible double cycles in n for this order volume according to 3.1 is

$$n = 3 \cdot 3 \cdot 2 \cdot 2 \cdot 1 \cdot 1 = 3! \cdot 3! = (3!)^2 = 36 \tag{3.1}$$

Thus, theoretically 36 combinations have to be analyzed with regard to the optimal sequence of trips. Since the time slot for all double cycles for the take-up and transfer of pallets remains unchanged the only optimization criterion is the minimization of the overall distance. By means of the distances[3] to be covered for fulfilment of the single orders in Fig. 3.1, the distances of the double cycles can be determined as well as the overall distance to be covered. (cf. Table 3.1).

[2] For a better understanding the storage locations to be approached and the transfer points are arranged in one level so that only the horizontal distances have to be considered when determining the optimal order of double cycles.

[3] The distances are entered without units into the coordinates of the storage location (in this case only the width).

It may be useful for the calculation that the order of double cycles seems to have no effect on the overall distance and that it does not matter whether the double cycle sequence is *1 - 5 - 9* (cf. Table 3.1) or *9 - 1 - 5*. As a whole this strategy reduces the expenditure for determining the best solution since only six sequences ($3 \cdot 2 \cdot 1 = 3!$) have to be built and compared.

In the present example the optimal order is the double cycle sequence *1-6, 2-5, 3-4*) with an overall distance of 38 bin units. The worst double cycle sequence is *2-4-9* (corresponding to the order sequence *1-5, 2-4, 3-6*) with 46 bin units. Generally, the following equation applies for the number n of possible double cycles at a symmetrical order volume of k storage and retrieval jobs each:

$$n = k! = k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1 \tag{3.2}$$

The problem described in equation 3.2 is characteristic for quite a number of similar optimization problems. A typical such sequencing or *scheduling* problem [4] is the *travelling salesman problem* (TSP), which is described in detail in the relevant literature and represents a basic optimization problem also in other technical and scientific areas.

A characteristic feature of the TSP is that the calculation expense grows unproportionately in line with the order volume and requires a large number of calculation steps and comparisons to find a solution. The present problem is relatively easy to solve *exactly* with six comparisons. It would require 10! = 3.628. 800 comparisons to find a suitable solution for ten double cycles. Assuming a calculation time of one second on a medium performance PC, it would take more than four days to search all solutions in order to calculate a sequence of 15 double cycles.

As described later in this chapter, despite increasing computer performances, the calculation time is a critical factor given the high number of orders which have to be considered when dispatching all work steps within a warehouse (cf section 3.3).

Multiple cycles at higher capacities

In containers or tray storage systems where load suspension devices with a higher capacity can be used, there are further possibilities to combine and thus to optimize transport trips.

The load suspension device of a rack feeder in a container store has a take-up capacity of two containers. The load suspension device should take up two containers at the transfer points TP1 and TP2 and store them at the locations 1 and 2 (cf. Fig. 3.2). In the course of this double cycle two other

[4] The present optimization problem provides for the determination of just one optimal sequencing of already known transport orders. Additional conditions such as the transport capacity of a rack feeder or the time slot for the order execution have not yet been considered.

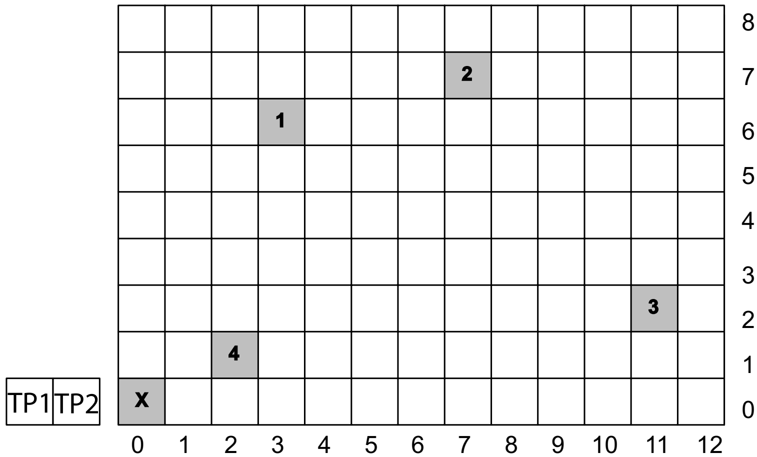


Figure 3.2. Example of a multiple cycle in a container store

containers have to be taken up at the locations 3 and 4 and transferred at the transfer points.

When the containers are stored and retrieved at random the combinations described in Fig. 3.3 are possible for the combined storage and retrieval of all four containers. Theoretically, the containers can be moved in one double cycle and thus in eight different trip sequences $(2 \cdot (2 + 1 + 1))$.

In order to find out which trip sequence achieves the shortest overall cycle time the times for all single trips have to be determined and added for each of the eight valid combinations.

In the present example it is assumed that the rack feeder, after having taken up the containers, carries out the first storage job from the reference point X where it will return after the last retrieval prior to the transfer. For reasons of simplicity, it is also assumed that the height and width of the shelves are equidistant and the horizontal or vertical speed of the rack feeder is identical. The trip time between two locations thus is proportionate to the longer section of a traverse path, i.e., the maximum of the horizontal or vertical section of a single trip.

Table 3.2 shows the possible eight trips and the resulting overall paths of the rack feeder. Compared to the best solution (26 unit lengths), the length of the worst solution (37 unit lengths) is 42% longer. This shows that trips can be optimized considerably optimizations also in small projects like this.

The sequencing problem occurs in many other fields of the material flow. Typical examples, other than the combined multiple cycles in container and tray storage systems, are the transport planning for stackers and tractors with a loading capacity of more than one pallet which should approach all storage

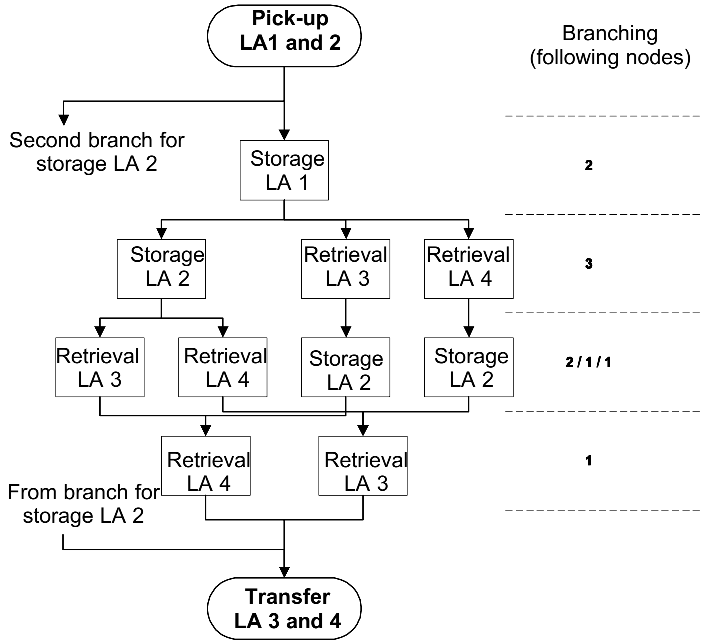


Figure 3.3. Trip sequences in a double cycle (load suspension device with double capacity)

locations on the shortest way. Similar problems arise during order-picking, according to the man-to-goods principle (cf. Fig. 2.2.5).

This problem, however, exceeds the TSP because the higher loading capacity of the conveyors increases the number of combinations considerably and thus the calculation input[5]. In the following, the calculation input for this optimization process, a *multi-capacity* TSP, is calculated on the basis of the *search area*, i.e., the number of possible order combinations.

A conveyor has a take-up capacity of n uniform loading units. The order should be executed in a closed cycle of known orders, i.e., no orders are added during the execution. At the beginning of the tour n loading units have been retrieved from the buffer and loaded on the conveyor according to its capacity which have to be stored at the respective sinks (E). Like in the example above n loading units have to be retrieved (A) and returned to the buffer. During this order cycle a total of $2 \cdot n$ loading units are moved.

Since all storage and retrieval jobs are known at the beginning of the tour the dispatching can be made in advance to determine the shortest tour for

[5] The example of a load suspension device shown in Fig. 3.3 with a double loading capacity allows for a higher number of possible sequences (eight combinations) than a load suspension device with a single capacity which executes the orders in double trips (2!=two combinations).

Table 3.2. Determination of an optimal frequency of multiple cycles

Sequence	**Single distances and overall path**
$X \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow X$	$6+4+5+9+2=26$, **shortest cycle**
$X \rightarrow 1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow X$	$6+4+6+9+11=36$
$X \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow X$	$6+8+5+6+2=27$
$X \rightarrow 1 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow X$	$6+5+6+5+11=33$
$X \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 4 \rightarrow X$	$7+4+8+9+2=30$
$X \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow 3 \rightarrow X$	$7+4+5+9+11=36$
$X \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 4 \rightarrow X$	$7+5+8+5+2=27$
$X \rightarrow 2 \rightarrow 4 \rightarrow 1 \rightarrow 3 \rightarrow X$	$7+6+5+8+11=37$, **longest cycle**

the approach of all loading points (sources and sinks) in the working area. During the sequencing, however, constraints have to be considered which are due to the limited capacity of the conveyors.

1. A tour has to start with a least one retrieval or storage (E) and
2. to end with the take-up or retrieval (A) of the last remaining unit load.
3. Only as many retrieval points (A) can be approached in succession as there are loading bins on the conveyor.

By this general task description all possible order combinations can be determined and based on the resulting search area the ones with the shortest overall path can be found. In order to describe the building of the search area orders may be shown as nodes in a *search tree* where the sequence of retrievals and storages is entered (cf. Fig. 3.4).

The figure shows a multi-capacity TSP with a capacity of two or three loading points. The nodes in the search tree each stand for an order within a sequence which has to be handled top to bottom. These nodes also bear information about the order type (storage E or retrieval A) with the number of variations for an order at this point.

Every complete branch describes the possible order sequence but only the *array* of successive storages and retrievals and not the definite order sequence[6].

An array stands for that order sequence where the storage and retrieval jobs are arranged in this way. The number of possible sequences [7] for individ-

[6] As an example there are the storage jobs E1, E2 and E3 as well as the retrieval jobs A1, A2 and A3. A definite (and valid) order sequence may then be E1-E2-E3-A1-A2-A3. The array is the qualitative order of storages and retrievals without clear information, in this case E-E-A-E-A-A.

[7] Invalid order sequences are, for example, those with an array E-<u>A-A</u>-E-E-A. The capacity of the conveyor would be exceeded by the first two take-ups.

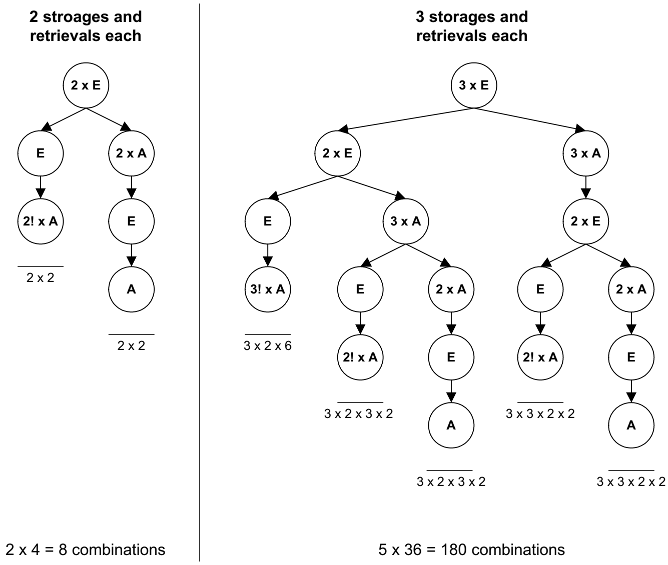


Figure 3.4. Search tree for multi-capacity TSP

ual branches can be determined by multiplying the coefficients of the nodes on a branch. The product is the same for all branches.

The valid order sequences for a conveyor capacity of k = 3 are shown in Fig. 3.4. In this case there are five branches with valid arrays. The topmost node may be occupied by one out of three possible storage jobs. The node coefficient for the top retrieval job is determined in the same way. The possibilities for the nodes further down decrease correspondingly. The number n_{ARRAY} of possible arrays at a given capacity k and the corresponding number of storage and retrieval jobs can be determined by the equation 3.3:

$$n_{ARRAY} = \binom{2\cdot k}{k} = \frac{(2\cdot k)!}{k!\cdot(2\cdot k - k)!} = \frac{(2\cdot k)!}{(k!)^2} \tag{3.3}$$

With a capacity $k = 3$ there are 20 possibilities for the sequencing of all storages and retrievals within a cycle. However, some are invalid because they are inexecutable.

The number n_{VALID} of valid arrays can easily be determined by means of the equation 3.4

$$n_{VALID} = \frac{1}{k+1} \cdot \binom{2\cdot k}{k} = \frac{1}{k+1} \cdot \frac{(2\cdot k)!}{(k!)^2} \tag{3.4}$$

The total number n_{TOTAL} of all valid order sequences can be determined as shown in Fig. 3.4

$$n_{TOTAL} = n_{VALID} \cdot (k!)^2 = \frac{(2\cdot k)!}{k+1} \tag{3.5}$$

Here, the term $(k!)^2$ represents the product of the node coefficient of a branch.

In this example, the total calculation input can be reduced by at least 75% compared to a complete solution which takes into consideration all possible (also the invalid) order sequences. This not only helps to determine the optimal order sequence but also the calculation input.

Considering the above-mentioned relations the volume of the sequencing problem can be reduced considerably. The strategy may also be applied to larger order volumes which are handled successively by a conveyor in several cycles.

As an example, four storage and retrieval jobs have to be carried out with a minimum input of a conveyor with a capacity of $k = 2$. Theoretically, the single orders can be executed in $8! = 40320$ different sequences. If the conveyor capacity is fully utilized and the orders are grouped so that two storages and retrievals are carried out in one cycle the number of possible sequences is reduced to $(4!)^2 \cdot 2^2 = 2304$.

This method is also used to reduce the search processes during an optimization. A general form of this procedure is the *Branch&Bound* (cf. section 3.3.3).

3.2.2 Sequencing of picking orders

The optimal sequencing of picking orders can also be described and solved as the "travelling salesman problem". Here, the retrieval points represent the locations, the pickers or the system technology represent the vehicles and the transfer/pick-up stations the depot. The optimal sequence of retrieval points has to be determined for each order.

Exact and heuristical methods have been developed to solve the TSP and tour planning problem (cf. section 3.3). Heuristics are used to determine the relevant problem because the runtime of the exact algorithm is too long. Tours can be planned with opening procedures which in the first step provide a valid but possibly sub-optimal solution. The solution may be optimized by subsequent improvement processes.

The TSP can be opened by finding the best successor or the strip strategy. At the best successor method the order with the shortest distance is chosen out of an order pool as successor of the starting order. Here, it is made use of the fact that the driving speed of a vehicle generally is higher than its lifting speed.

For the strip strategy, the cross-section of the warehouse is divided into strips of the same size. The shelves in the single strips are approached in turn. At the end of each trip, the picker goes to the next strip. This procedure can be improved by swapping the edges, e.g., with the 2-opt or the 3-opt method where the solution should be improved by swapping the edges in the respective graphs.

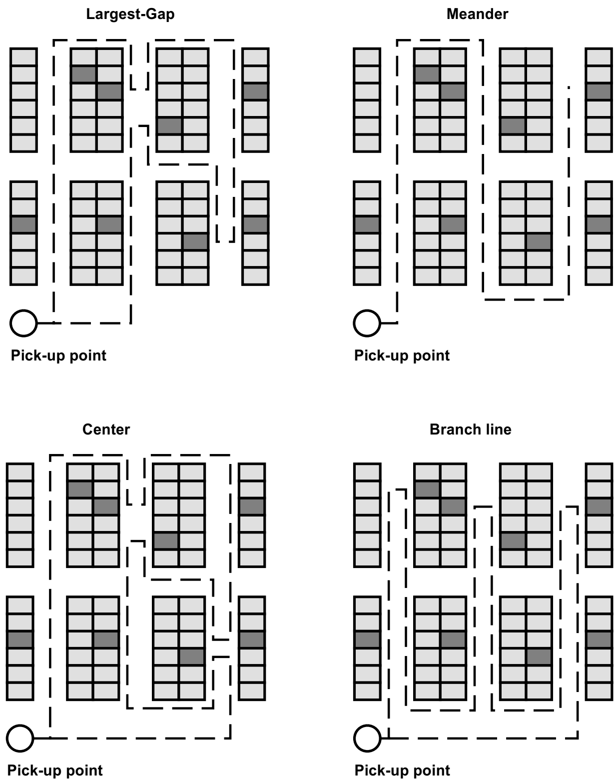


Figure 3.5. Order-picking strategies

Heuristic tour planning methods (cf. Fig. 3.5) are differentiated to whether the aisles are passed (meander heuristics) or if the picker turns around in the current aisle (center or largest gap heuristics). At the meander heuristics the picker passes an aisle if it contains at least one target shelf. The aisles are passed in a predetermined order either left to right or the other way round.

If more than one shelf has to be approached in one aisle the best alternative is the center heuristics. Here, each aisle is split in two parts. The shelves in the upper half are approached from the upper aisle, the shelves in the lower part from the lower aisle. The picker leaves the aisle at the side where he has entered it.

Another excellent method is the largest gap heuristics. Unlike at the center heuristics the aisle is passed up to the largest gap. A gap is the distance between two neighbouring shelves to be approached or the distance from the aisle to the next shelf. The largest gap is that part of the aisle which is not passed.

3.2.3 Routing in the warehouse

The purpose of routing (also trip optimization) in the warehouse is to determine the shortest way if there are several possibilities. Unlike in a high-bay warehouse where the storage locations are approached with the shortest way from the storage point the schedule of a driverless transport system, for example, may allow for different routes for one tour. In some warehouses, transports with stacker cranes offer several possibilities [8].

A typical situation is shown in the figure below where a consignment has to be moved from the given source to the marked sink.

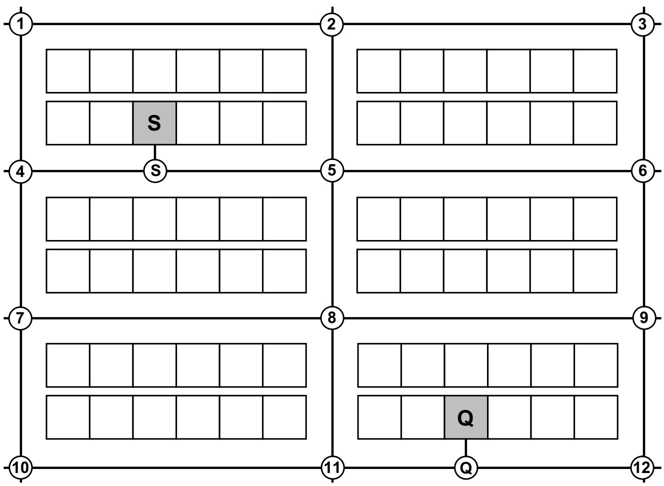


Figure 3.6. Routing problem in a warehouse

The warehouse layout consists of storage locations arranged in rows which are connected by an orthogonal network of ways. The shortest way from the source between nodes 11 and 12 to the sink between nodes 4 and 5 is approached on the route → 11 → 8 → 5 → S. Principally, however, all other possible (and longer) routes between both final nodes of the tour can be chosen.

In order dispatching a computer-aided vehicle control system is used to determined the shortest route for the execution of an order. In addition to a path matrix this step also requires information about the path network in the warehouse to be able to determine the shortest way. One method to determine the shortest way is the *Dijkstra* algorithm, by which the way between a starting node and a target note can be determined in a graph (cf. section 3.3.3).

[8] In these cases, it depends on the ability of the staff to find the shortest way to the sink.

3.2.4 Comprehensive order dispatching

The goal of an order-related batch generation is to manage orders which have been received until a certain deadline in fixed intervals, so-called *batches*. The term *batch calculation* is used in different contexts and describes the prearrangement of picking or sorting orders as well as the complete order management in a warehouse.

Within a batch all received orders have to be arranged and assigned to the warehouse resources as to achieve the shortest possible processing time in line with a continuous system load or utilization of resources. A general term for this kind of problem is *job-shop-scheduling*.

The following examples describe the batch scheduling for a warehouse [78]. A superordinate ERP system transmits jobs to a WMS where customer orders are accepted, delivery dates are coordinated and preprocesses according to the current stocks and the job load. An order transmitted to the WMS generally consists of the following elements:

– Order number (clear)
– Customer number
– Number of articles to be delivered for this order
– Date for the provision at the goods issue (and then collection by the shipper)

The order intervals are an important parameter for the batch planning since the optimization potential is the largest for large order volumes. The scheduling of order-related delivery dates stands in direct connection to the batch scheduling and has to be determined based on the current resource productivity values. In some cases this kind of dispatching is impossible: an *exact* forecast of a delivery date would imply that this order had to be integrated into the current backlog. The orders of this backlog, however, are also given time intervals for delivery so that it would be necessary to recalculate the order schedule[9].

An order dispatching method used in many companies is based on the medium-term estimation of the productivity so that articles ordered up to a certain time can be provided for pick-up at the goods issue until the next day. The order dispatching in the WMS (and thus the batch scheduling) is typically carried out as follows:

– Acceptance of orders by the host system
– Separation of orders into single items and multiple item jobs
– Determination of storage locations for single articles
– Determination of the required number of containers for picking
– Calculation of the handling time (lead time) of each job

[9] An exact batch dispatching method is described in [3]. According to this method it takes several minutes to calculate the daily picking schedule in a picking area with some thousands of items by means of an exact method (CSP).

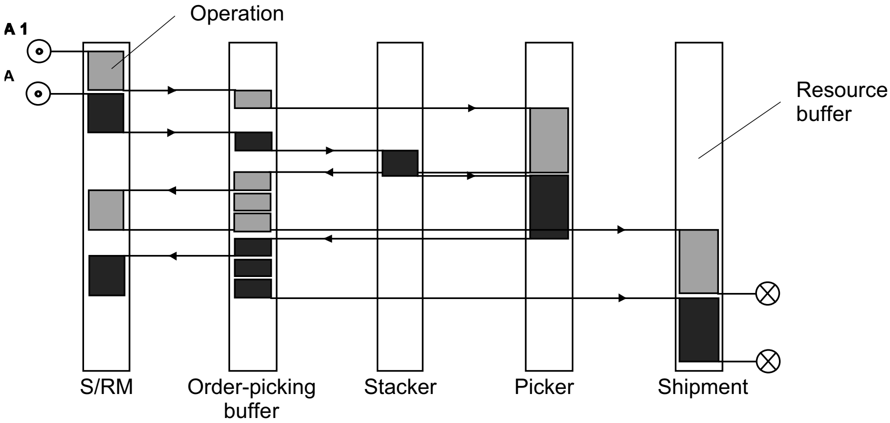


Figure 3.7. Assignment of orders to resources presented in the form of a Gantt diagram

The order lead time consists of single steps, called *operations*, e.g.:

- Picking time (basic time, gripping time, idle time, way time; cf. section 2.2.5)
- Provision of empty loading facilities
- Buffering after the picking (depending on whether goods are picked directly before shipment, working hours, shift model)
- Provision at goods issue

Multiple item orders consist of similar operations which have to be sequenced optimally already in this planning step, e.g., in case of discrete pickings. The operations have to be assigned within the batch planning to the warehouse resources.

Thus, all customer orders are separated into elementary operations, which are scheduled backwards starting with the respective delivery dates. Each resource is assigned to an order pool which can be locally optimized later on (cf. Fig. 3.7).

The sequence of orders in a pool can still be modified if the time schedule (valid at that time) allows for such a change.

The time needed to plan the batches mainly depends on the number of orders to be planned and thus on the working time to be represented (time intervals). In principle, the batch size or the time interval can be detailed at random so that the calculation time is reduced considerably. It has to be considered, however, that smaller optimizations are achieved with short time intervals because the resources can be fully occupied for this complete interval to a limited extent.

On the other hand, when smaller intervals are used more information are available about the occupancy of resources because orders can be dispatched and batches can be recalculated at any time. The batch scheduling in a WMS has to fulfill these two requirements — optimized order throughput and short-term dispatchability.

3.3 Optimization of solutions

3.3.1 General aspects

The problems which arise during the operational optimization and the dispatching are *computable* i.e., they can be described by an algorithm.

The *complexity* defines the computation time of an algorithm and thus is of vital importance for its development and use. Therefore, it is also called computational complexity. Based on the algorithm structure an estimation is made to determine the complexity. This estimation reveals the demand for resources [10] of an algorithm at a given problem volume n.

The time complexity $T(n)$ can directly be determined as a function of the n input data by counting the operations computed by the algorithm. Since emphasis is given to the functional behavior of n rather than to exact digital values the time complexity is represented by the $\mathcal{O}$ notation which represents the highest degree of that polynomial which describes the runtime of the algorithm. As an example the time complexities of a typical task are shown in ascending order:

- $\mathcal{O}(log\ n)$: Logarithmic expenditure, e.g., binary search in n structured data
- $\mathcal{O}(n)$: Linear expenditure e.g., addition of n input values
- $\mathcal{O}(n\ log\ n)$: quasi-linear expenditure e.g., sorting
- $\mathcal{O}(n^2)$: Quadratic expenditure e.g., vector multiplication
- $\mathcal{O}(n^k)$: Polynomial expenditure e.g., matrix multiplication
- $\mathcal{O}(2^n)$: Exponential expenditure e.g., accomplishment problems
- $\mathcal{O}(n^n)$: E.g. travelling salesman problem (TSP)[11]

$\mathbb{P}$ describes the category of problems which can be solved with an *deterministic*[12] algorithm in a *p*olynomial and thus practically realistic time, i.e., $\mathcal{O}(1)$, $\mathcal{O}(n)$ or $\mathcal{O}(n^k)$. $\mathbb{NP}$ describes the category of problems which can only be solved by a *non-deterministic* algorithm in a polynomial period of time.

[10] Generalized for disk space and time needed to find a solution. The complexity is a combination of the time complexity $T(n)$ and the memory complexity $S(n)$.

[11] An algorithm for the complete solution of the TSP is exactly described by the runtime behavior $T(n) = (n!)$ while the order $\mathcal{O}$ is determined by means of the Sterling formula $n! \approx n^n \cdot e^{-n} \cdot \sqrt{2\pi n} \approx n^n = \mathcal{O}(n^n)$.

[12] An algorithm which generates an (optimal) solution after a predetermined period of time.

According to the current state of computer sciences a deterministic algorithm needs a time which exponentially increases the size of the problem.

In addition to the formal definition the following features are characteristic for an algorithm:

- *Finiteness* – A time-phased algorithm solved in a problem-based, predetermined number of computing steps.
- *Clearness* – The further steps are determined at any time.
- *Effectiveness* – Each work step in carried out in a finite period of time
- *Efficiency* — The algorithm requires minimum disk space and computing time.

The algorithm is always closely related to the predictability of the *solution quality* while the term heuristics[13] is used for those procedures where the quality of the solution cannot be predicted.

3.3.2 Overview over the optimization procedures

A variety of different methods is used to optimize combined problems. These methods can be differentiated by the descriptive algorithm or heuristics, the runtime behavior and their robustness. The following classification shows the basic procedures while a concrete implementation is described in the next section.

Assignment strategies in queues

Assignment strategies are an efficient method to assign orders to resources and are mainly used to solve *online problems*. The order pool of a resource has to be considered as a queue from where the orders are retrieved successively. New orders can be added to this queue at any time (Fig. 3.8).

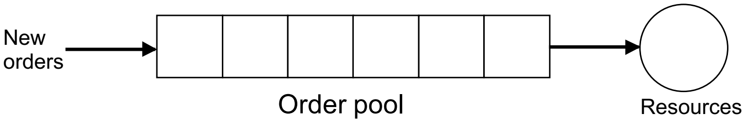


Figure 3.8. Dispatching in queues

The optimal sequence of all orders in the queue can be calculated only to a limited extent since the combination of orders varies over the time. Thus, the principle of the "best successor" is used for these cases. Such *dispatching strategies* pick an order out of an order pool of the size n, which changes over the time with a linear searching expenditure of $(\mathcal{O}(n))$.

[13] Greek *heuriskein*: to find.

An optimization, i.e., the search for the best successive order, focusses on application-specific criteria. For conveyors and order-picking systems this may be the *shortest connection trip*. Other criteria are the priority of the order or the dispatching date. Search dispatching strategies are studied in [3, 64].

Enumerating methods

In completely enumerating optimization methods all solutions within a scope of solutions are studied so that it is guaranteed that the exact *optimal* solution is found. However, all solutions have to be generated and evaluated successively with the respective expenditure.

Since the number of solutions for real problems may be enormous *completely enumerating* methods are in this case not very efficient. The typical sequencing and scheduling problems belong to the group of $\mathbb{NP}$ complete methods with exponential solution expenditure.

Adapted enumerating optimization methods reduce the solution expenditure by integrating heuristics. The complete enumerating methods are modified in that way that a minimum of solutions have to be generated and evaluated while at the same time generation of the best solution (global optimum) is ensured. A known example for such a method is the *Branch&Bound* algorithm.

Calculation-based methods

With *indirect* calculation-based optimization methods the optimal solution is found by means of a gradient method which studies the target function in dependence of the parameters. Based on the target function the gradients are set at zero and then the resulting equation for the corresponding parameter is solved. The results represent the local optima which are compared and evaluated. With this method the global optima can be identified.

With direct calculation-based optimization methods the best solution is searched by targeted steps within a search area. A randomly chosen starting point represents an initial solution. Near this solution the direction with the steepest incline, i.e., the highest gradient value is determined.

The method searches the solution scope iteratively and stops when there are no better solutions in any direction. This method is called “hill-climbing method” because it unerringly finds the nearest local optimum.

In the best case the last local optimum is also the global optimum. But since the global optimum is not guaranteed the result is just an approximation. It has to be taken into consideration that only the enumerating optimization method ensures the best solution although calculation-based methods in most cases require considerably less expenditure.

Random methods

Random or *stochastic* optimization methods operate without a targeted search but iteratively generate random solutions which are evaluated and improved until a stop criterion is reached.

With the *Monte-Carlo strategy* different random solutions are generated and evaluated. If a better solution is found, it overrides the last one. A common stop criterion is the number of iterations.

Random walk is another method where the solution scope is searched step by step based on the initial solution. Each solution is developed by marginally changing the last one at random. The best solution is saved and the stop criterion may also be the number of generated solutions.

Nature-analogous optimization methods also belong to the group of stochastic methods. Analogous to natural processes the chemical/physical and biological models are differentiated.

Conclusions about the quality of the solutions cannot be made neither by means of the Monte-Carlo strategy or the random walk nor with one of the nature-analogous methods. In some cases not even a local optimum is found.

3.3.3 Examples of known methods

Some highly specialized methods have been developed to solve all kinds of optimization problems. The algorithms described below represent some of the published methods for the optimization of warehousing problems.

Branch&Bound algorithm

The Branch&Bound principle is an exact method where systematically promising solutions are generated based on the solution of simple subproblems and sub-solutions which do not result in the optimum are excluded.

In the following example the shortest route shall be determined to demonstrate the functioning of the Branch&Bound algorithms (cf. Fig. 3.9). The respective graph shows the knots 1 to 5 and the drawn side lengths. Node 1 is the starting node and node 5 the target node of a certain tour.

To point out the single route finding steps the solution is described in 3.10 of the diagram.

Distance, **predecessor** and **lower limit** each represent a field the size of which corresponds to the number of nodes in the graph. Like with the Dijkstra algorithm the next known **Min** is taken from a quantity **B** of nodes still to be processed whereas **B** corresponds to those nodes in the graph which can be reached directly from the already studied nodes.

Starting from this node **Min** all adjacent nodes are studied. If the neigbouring node still has no predecessor **Min** is entered as **predecessor** and the path length over **Min** to the starting node is entered as **distance**. If there is a predecessor this is replaced by **Min** if the distance from the starting node is shorter via **Min** than the distance saved for this neighbour.

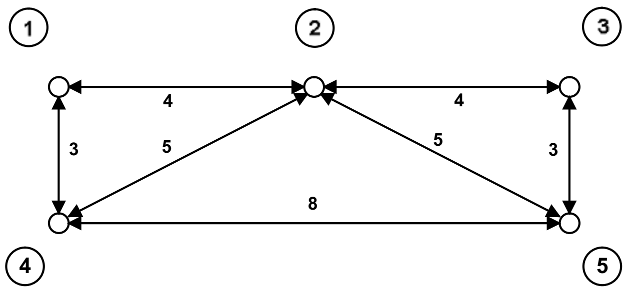


Figure 3.9. Example of a TSP

B	Min	Distance 1	Distance 2	Distance 3	Distance 4	Distance 5	Predecessor 1	Predecessor 2	Predecessor 3	Predecessor 4	Predecessor 5	Lower limit 1	Lower limit 2	Lower limit 3	Lower limit 4	Lower limit 5
1		0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2 4	1	0	4	0	3	0	1	1	0	1	0	0	9	0	11	0
3 4 5	2	0	4	8	3	9	1	1	2	1	2	0	9	11	11	9
	5															

Figure 3.10. Determination the shortest route

As can be seen in fig. 3.10 in line 2 in the example the estimated path lengths for the adjacent nodes of the starting node 1 are calculated as follows:

$$
\begin{aligned}
Lowerlimit(2) &= 0 + 4 + 5 = 9 \\
Lowerlimit(3) &= 0 + 4 + 4 + 3 = 11 \\
Lowerlimit(4) &= 0 + 3 + 8 = 11
\end{aligned}
$$

The most promising node, thus, is node 2 which is therefore taken from the quantity **B** to study its neighbour. The search for the best route is terminated when the target node is taken from the quantity **B**. In this example this happens when node 5 is reached after only 3 steps. The shortest route can then be determined with the field *predecessor (1-2-5)* and its length can be derived in **distance** (5) = 9.

The *Dijkstra* algorithm determines the shortest route by means of a graph is also based on this method.

Heuristic openings

The goal of an opening is to determine valid order sequences which do not necessarily have to be optimal because they are improved later on by a respective method. The following method describes the solution of a TSP with n locations and one starting point similar to the computation of an order-picking sequence. It is the basic idea of the *savings heuristics* to approach all locations individually directly from the starting point (0). Such a solution consisting of commuting tours ($0 \rightarrow$ j $\rightarrow 0$) seems to be rather unfavourable because the overall distance resulting from the sum of all single distances d_{0j} (distance between 0 and the location j) would be computed very inefficiently with

$$2 \cdot \sum_{j=1}^{n} d_{0j} \tag{3.6}$$

But if a tour is chosen in such a way that two location s_{ij} are approached in one cycle savings s_{ij} can be achieved by

$$s_{ij} = d_{0i} + d_{0j} - d_{ij} \tag{3.7}$$

These "savings" are computed for all location pairs which allow for a joint tour under consideration of possible restrictions. The savings algorithm can be formulated as follows using a distance matrix $D = d_{ij}, 1 \leq i, j \leq n$ [108]:

1. Generation of an initial solution, all commuting tours $(0 \rightarrow j \rightarrow 0)$ for $j = 1, \ldots, n$.
2. Determination of all savings according to eq. 3.7 which may be generated under consideration of the restrictions. Sorting of computed savings s_{ij} in descending order in a list.
3. Search for an edge d_{ij}, for which s_{ij} is the maximum. Connecting the locations i and j if the following applies:
   – The locations are not on the same route.
   – None of the locations i and j is an internal point of a route.
   – No side conditions (capacity, time, driving distance, etc.) are violated.
   Cancel s_{ij}, continue with step 4.
4. Check whether more tours can be combined. If so, continue with step 3, otherwise step 5.
5. End

Other openings are "best successor", the sweep method [107], especially for the planning of picking orders and the methods described in 3.2.2.

Deterministic r-opt improvement methods

The *r-optimal* methods assume a valid solution determined by an opening and try to improve the quality by exchanging a number r of edges in a sequence graph. In 2-opt methods the possible exchange of two edges is studied. If this improves the solution the edges are exchanged.

The algorithm for the 2-opt method to determine the shortest tour in the scope of a TSP is described as metacode. Known factors are the n locations s_1 to s_n and the distance matrix C where all distances between two locations s_i and s_j are entered as a way distance c_{ij} as well as an allowed starting tour $r = [t_1, t_2, \ldots, t_n]$ with $t_{n+1} = t_1$. The algorithm is terminated if no improvement can be achieved in an iteration.

```
ITERATION μ (1,2,...)
    For i := 1 to (n-2) do
    begin
        For j : = i + 2 to n do
        begin calculate △c_{t_i t_j} + c_{t_{i+1} t_{j+1}} − c_{t_1 t_{i+1}} + c_{t_j t_{j+1}}
            If △ < 0 generate a new tour
            [t_1, ..., tn] with exchanged t_i and t_j
            Continue with ITERATION μ + 1
        end
    end
```

Genetic algorithms

Genetic algorithms (GA) belong to the group of nature-analogous methods which adapt the mechanisms of natural systems in the form of technical and economical applications [32]. The idea of developing and using genetic algorithms to solve technical optimization problems was initiated by observations and the wish to imitate nature which follows the principle of the "survival of the fittest" [14].

Genetic algorithms generate better (and some times worse) solutions out of a pool of initial solutions within the *population* by means of *recombination*, i.e., a new combination of already existing solutions. The worst solutions are excluded from the pool and replaced by better ones until no significant improvements can be achieved after several iterations. To ensure that the GA produces no identical solutions the *mutation* method is used, i.e., a recombined solution is changed at random.

The following steps describe the basic procedure of a GA. It assumes a problem size divided by the number of elements n for which an optimal sequence should be determined:

1. Generation of a basic population of μ elements (initialization, $\mu \ll n!$).
2. Definition of a stop criterion for the following closed loop.
3. Evaluation of the individuals with the target function f (evaluation of the fitness by means of the path length),
4. Choice of individuals from this population (selection depending on the fitness)
5. Generation of offspring (recombination)
6. Modification of the characteristics of the offspring (mutation).
7. Repetition beginning with step 3 until the stop criterion is reached.

[14] According to the evolutionary theory the "Survival of the Fittest" is the result of selection, inheritance and natural random mechanisms, the mutations which lead to individual changes and characteristics of biological organisms.

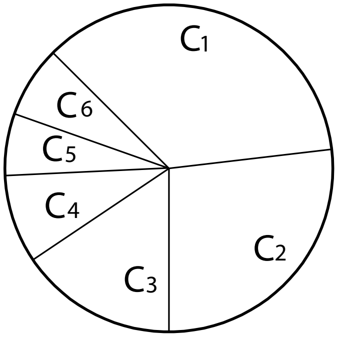


Figure 3.11. Selection according to the fitness ratio (roulette wheel)

Sequencing problems are encoded as GA in such a way that the orders to be sequenced each correspond to one *gene* within one strand the *genome* or *chromosome*. Each order may take any place within the chromosome and the execution sequence will then correspond to the position within the chromosome.

During the *initialization* a random generator generates a basic population out of μ different chromosomes. With regard to sequencing problems the fitness of these chromosomes corresponds to the tour length determined in an evaluation (e.g., order-picking tour[15]).

In the scope of a selection the parent chromosomes are chosen for recombination according to their fitness. Based on the relative frequency of the fitness a simple choice is made (cf. Fig. 3.11). But this may lead to *dominance effects*. Certain chromosome characteristics would be selected with an overproportionate frequency. Therefore the GA just focusses on a local optimum. This effect can be avoided by *normalising* the fitness values.

After two parents have been selected, their genes are bequeathed to their offspring by recombination operators. An operator for sequencing problems which is easy to implement, in addition to various problem-specific operators, is the *uniform order-based crossover* where a random bitstring along the chromosomes determines which genes will be bequeathed to which offspring (Fig. 3.12).

The points in the bitstring marked with 1 bequest the genes of parent 1 to the first offspring while the other genes marked with 0 are bequeathed by parent 2 to the second offspring. The empty spots in the chromosomes of the offspring are filled with the parent genes corresponding to their order in the parent genomes. Thus it is ensured that only valid genetic sequences are generated for the offspring (no doubling of genes).

[15] It has to be considered that the fitness is maximized or minimized according to the problem, i.e., depending on the basic target function.

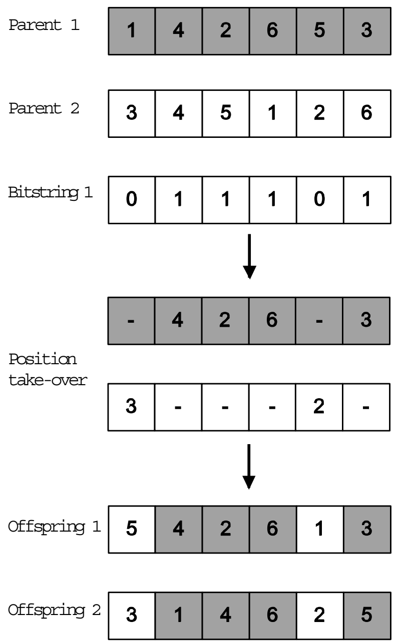


Figure 3.12. Recombination operator *uniform order-based crossover*

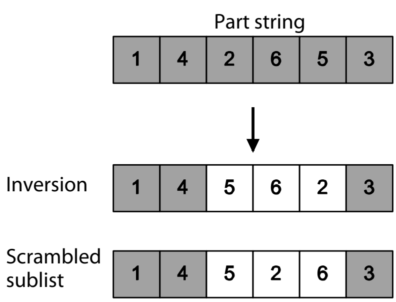


Figure 3.13. Application of mutation operators

By *mutation* single chromosomes within a population can be modified purposefully to limit the scope of solutions and thus to avoid a mere local optimization. In case of the mutation operators shown in Fig. 3.13 sections of a chromosome are arranged either by mirroring the gene sequence (inversion) or by a random rearrangement (scrambled sublist).

How the different operators are used for reproduction depends on the *reproduction model* of the population or GA. This model defines, among others, the number and possibility p_{cross} of the parental pairs to be selected and recombined and their offspring as well as the number of generations as well as the respective mutations and clonings with the possibility of their occurrence.

At the same time the reproduction model classifies the computation expenditure and thus is a measure for the long-time behavior. Generally, the computation expenditure for the GA described above can be considered as polynomial.

In contrast to the classical optimization method, a genetic algorithm does not rely on the stepwise improvement of just one solution but on several solution instances which occupy different places in the search area at the same time. Thus, genetic methods are more resistant to changes of the input values which determine the problem and can, thus, ensure that the solutions are generated within a locally limited area (local optimum). For a further study of genetic algorithms please refer to [36, 43].

4. Warehousing and Conveying Principles

Corresponding to the variety of problems and tasks concerning the storage and transport of piece goods, there is a similar multitude of systems for their effective solution. In the following chapter the most common technical solutions for the storage and transport of goods are presented as a basis for a study of the efficient control and management by means of a warehouse management system. The intention of this chapter is not to compare different systems and technologies with the aim to optimize the choice of technologies or systems but to provide a basic knowledge about these systems. For more comprehensive and comparative studies please refer to other publications [40, 54, 72, 73].

4.1 Warehouse systems

The correct evaluation and choice of warehouse systems calls for a systematic and general knowledge of the system performance. The basic features of different warehouse systems are shown in Table 4.1 as well as some general characteristics. Warehouse systems are classified into statical and dynamical systems depending on whether they require a relocation during the storage process. The restorage in a classical rack system is not considered as a dynamical storage.

Some basic characteristics can be identified by this classification. The primary parameters for the choice of a storage system are:

- Number of different articles
- Article dimensions and weights
- Quantity of each article
- Required storage/retrieval performance or throughput
- Space requirement
- Retrieval behavior and strategies

The most commonly used systems are described below and classified with regard to their applications and requirements on the system control. They can further be systematized according to the storage function, the type of construction and the used conveyors.

Table 4.1. Differentiation of warehouse systems

Characteristic	Type	Description	Common targets
Technology	Ground storage	Goods are stored directly on the floor and stacked, if necessary.	Inexpensive storage of large quantities of a few articles.
	Shelf storage	Goods are stored in shelves, mostly by means of a loading aid.	Direct access to large number of articles. High space utilization.
Type	Block storage	Goods are stored directly on top of each other, side by side and one after the other.	High utilization of space and short operating distances.
	Line storage	Goods are stored directly on top of each other, side by side and one after the other; there are operating aisles between the racks.	Direct access to a large number of articles.
Location	Statical storage	Goods remain at the same location from storage until removal, i.e., they are not restored.	Inexpensive storage technology. Little stress onto the stored goods.
	Dynamical storage	Goods are moved after the storage; although storage/retrieval at the same location is possible.	Short operating distances. Direct access despite of a high volume utilization.

4.1.1 Ground store

The goods are stored or stacked directly on the floor. The stacking height depends on the characteristics of the goods or on the used loading aids (e.g., skeleton boxes), the handling facilities (e.g., stacker or crane) and on the construction layout.

This simple form of storage requires only little investment costs and can be adapted flexibly to the constructional conditions (layout and form of the building). When the aisles are wide enough, relatively high transshipment performance can be achieved with a corresponding number of transporters. This type of warehouse is operated mainly manually.

Ground block storage The goods are arranged in a compact block, i.e., directly one upon another, side by side and one after the other. The space is

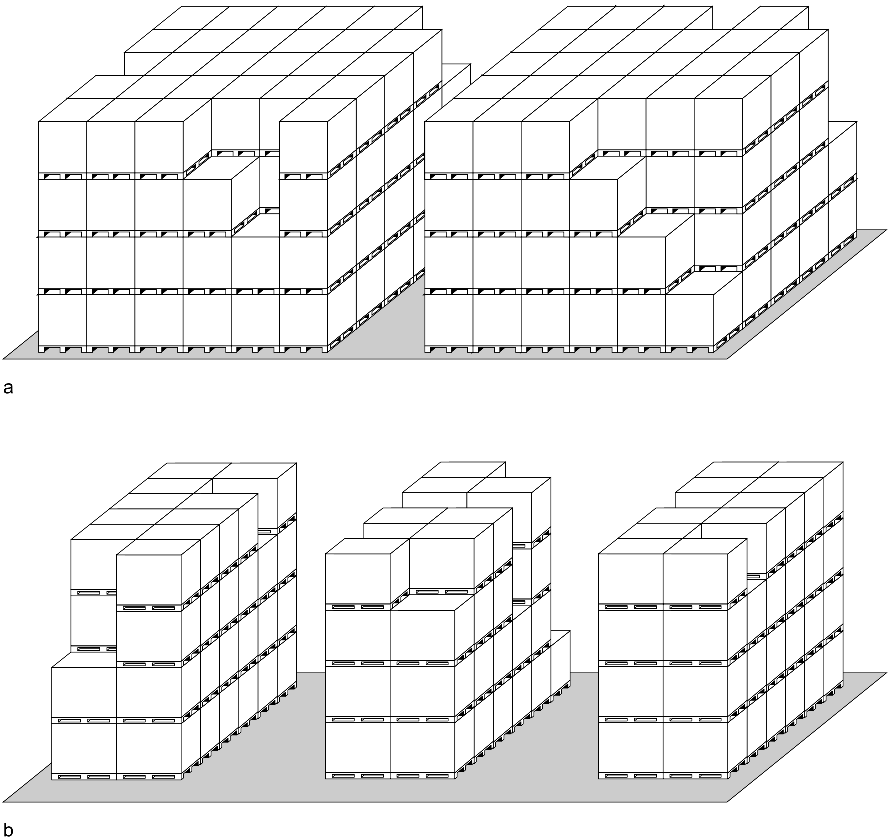


Figure 4.1. Ground storage a) Block storage b) Line storage

optimally utilized by this method, but only the goods in the first row can be accessed. Therefore, this principle can be used only in warehouses working according to the LIFO strategy (cf. Table 2.4, p. 29) These may either be classical mono-structured warehouses (beverages, raw materials) or buffer warehouses in the goods receipt/issue if complete loads are buffered.

Ground line storage In order to give better access on the single unit loads, compared to the ground block storage, the articles are arranged in such a way that each column is located at an aisle. This reduces the degree of space utilization but increases the number of purposefully stored articles.

Organization and operation The disadvantage of this simple system is the difficult control of the material flow. Since only ground space is recorded as the storage location and the storage and retrieval is controlled manually and according to the article type, it is not possible to record a single unit load.

The same applies to the direct assignment of a storage location to an operator (stacker driver) since precise reference points such as the shelf numbers on a rack are missing. There is no exact bin management, quantities are managed by recording the incoming and outgoing units.

This situation possibly leads to problems with the tracking and tracing of batches. If such a tracking is required single batches can be stored in separate fields. In more advanced systems the exact position is recorded during the storage and retrieval process by GPS (global positioning system) or based on the AGV (automated guided vehicle) technology to ensure the tracking of the material flow [60].

On the other hand, in fully automatic block warehouses which are operated by cranes or automated guided vehicles (e.g., for the storage of paper rolls or steel coils), the single items are recorded and managed exactly without the problems mentioned above.

4.1.2 Statical racking systems

Racks are mostly used to optimally utilize the space and height of a warehouse. The unit loads are placed on a separate shelf or a specified bin in a rack. This allows for the efficient storage also of not stackable goods. The rack height is mainly determined by the used operating facilities (cf. section 4.2). Accordingly, the rack heights range from 2 m (manual operation) to about 50 m (for rack feeders).

In addition to the problems of the physical storage, questions arise with regard to the choice of a racking system. A great advantage of this system is the possibility to clearly assign unit loads to storage locations and the almost complete implementation of necessary strategies. This sets the basis for the automation of storage processes and ensures an that the storage area is ordered and well-organized during the daily operation.

Goods stored in line racks can be accessed from any side while block racks allow for a more compact storage and a high utilization of space in line with a high throughput.

Most racking systems, however, are suitable only for single-type goods and standardized loading aids. If only complete unit loads are stored and retrieved they are also called *unit warehouses*.

Line racks

The design of a line rack varies according to the size, form and weight of the goods to be stored, the chosen or possible handling technology, the required storage and retrieval performance and the constructional conditions. The single shelves have to be adapted to the maximum dimensions of the goods to be stored plus all handling and transfer tolerances. The length of the individual lanes and the arrangement of the rack aisles are mainly determined by the requirements of the order-picking system.

The term "line rack" means that single shelves are arranged on top of each other and side by side where the unit loads can be stored and retrieved directly at the front. In case of a normal single-deep storage each unit, the load can be accessed directly so that all possible storage strategies can directly be implemented (section 2.2.3). Strategies like FIFO or LIFO are based on the warehouse organization and not on the physical arrangement of storage locations. Furthermore, the unit loads can be stored *double* or *triple deep* with a suitable handling technology. When unit loads in the back row have to be accessed, however, this requires restorage or special storage operations which reduce the throughput[1].

Provided the system has a functioning bin management all unit loads can easily and quickly be localized and retrieved. Line racks are always of advantage where moderate quantities of a large variety of single articles have to be stored.

Different goods require different types of construction which are called pallet rack, container rack, box rack or box warehouse depending on the type of loading aid or traverse rack, cantilever rack, honeycomb rack according to the construction of the warehouse. These systems are further divided into storage systems with and without loading aids.

Large and heavy unit loads are generally handled with fork lift trucks, rack feeders or cranes which vertically lift the units. In some cases also satellite vehicles are used which are equipped with load suspension devices for a sidewise take-up (automatic distribution vehicles). Light units are generally handled manually. The width of the aisles and thus the space utilization varies depending on the used handling technology.

Pallet rack Pallet racks are the most common type of pallet storage systems and are designed for use of a standardized loading aid. In the classical case the stored unit load (pallet or skeleton box) is supported only at both front ends whereas the unit load is stored lengthwise or crosswise. In case of a lengthwise storage two traverses are fixed (screwed, fitted or welded) between the front and rear rack supports where the unit loads are stored side by side. Due to their construction standard pallets (800 mm×1200 mm, 1000 mm×1200 mm) have to be stored lengthwise referred to the shelf depth (cf. Fig. 4.2). Because several unit loads may be stored side by side this principle is considered a *multi-bin storage system*. Generally, three up to max. five unit loads are stored side by side in a so-called *field*.

For a crosswise storage an angular support is mounted between a front and rear support and the unit load is stored frontal to the support and thus crosswise. There is only one unit load between the rack supports so that this system is a *single bin storage system*.

[1] Actually, strictly speaking the multiple-depth storage is a block storage. But as most systems use a double-deep storage and multi-functional load suspension devices these systems are considered as line racks.

Figure 4.2. Pallet rack with lengthwise storage [Photo: STÖCKLIN SIEMAG]

The lengthwise storage allows for a more efficient utilization of storage space since the racks require less supports and security distances. Most rack feeders are not so small as to allow for narrower aisles. In manual order-picking, however, the crosswise storage ensures a better accessibility of the shelved articles (the maximum reach of a person is about 950 mm).

In case of a double-deep storage the heights of the traverses have to be adapted to the constructional height and the bending of the load suspension device.

If pallets with different dimensions are stored which do not fit the traverse grids, sheets or boards may be placed on the traverses so that, strictly speaking, a kind of shelving systems is created.

Container racks When very small articles or small quantities have to be stored the standard pallet size (800 mm×1200 mm) often does not lead to a satisfactory utilization of space. For this reason smaller containers or trays are used for this purpose (400 mm×600 mm). Trays are metal troughs with a gripping bar at the front. Since the articles have a low weight they can be stored on simple angular profiles which are mounted at the sides of the shelves.

Figure 4.3. Automatic mini-load system with trays [Photo: BITO]

Thanks to the low weight the shelves can be served more efficiently what resulted in special designs of line racks and are called *container, box* or *tray racks*. Owing to the relatively low weight of the items the unit load (container, box, tray) can be pushed into or pulled out of the shelf. For this purpose the load suspension device grips the bar or handle of the tray or removes the unit by means of gripping pliers (cf. Fig. 4.24). With such pulling technologies shorter load transfer times can be achieved and the vertical safety can be reduced so that the storage space can be utilized more efficiently.

The container racks are generally served by automatic rack feeders and called *automatic mini-load systems* (cf. Fig. 4.3). If mini-load systems are used in accessible areas, they have to be encased at the sides to avoid a shifting of the unit loads which may lead to considerable damage in case of a collision with a rack feeder. Special attention is also given to operator and theft protection.

Mixed pallet storage The storage of different single item containers on one pallet or *mixed pallet* puts special requirements on the warehouse management. In addition to the management of several units at an identical location (what is generally done by booking the units on one pallet ID), the quantities stored on the pallet have to be controlled because the retrieval frequencies differ for each unit. This may even require consolidations (cf. section 2.3.2). Furthermore, if an article has to be added to a partially loaded pallet this unit has to be retrieved what causes additional ways. This problem also arises in connection with mixed storage on trays and in containers.

High-bay racking The term *high-bay warehouse* is often used to describe high racking systems although it is correct only for a rack height from 12 m onwards. The term *high-bay warehouse* describes a high-bay racking system with fixed storage and retrieval machines (S/R). High-bay racks can have a height of up to 50 m and may offer space for more than 100 000 pallet bins or several 100 000 container bins.

Such systems are often constructed as silo systems where the racks support the ceiling and the walls and thus build a single-purpose construction which is only used for storage purposes. One advantage of this system, in addition to the short implementation time and relatively low costs, is the considerably shorter amortization time. However, the rack construction has to allow for the storage of all other kinds of loads.

Lift systems (tower racks) A special type of line rack is the lift system or tower rack. Here, two storage columns are arranged directly face to face. In between drives a vertically operating special load suspension device which pulls the trays and moves them between the shelves and the transfer station (cf. Fig. 4.4). Owing to the low weights, high acceleration and thus high storage and retrieval performances can be achieved.

In addition to systems with fixed shelf heights in a storage column, systems with flexibly definable bin heights are used. For this purpose, other than fixed shelves a take-up grid for trays is created with a grid width of about 100 mm where the trays are inserted. After the height of the unit load has been recorded the tray is stored and the corresponding level is blocked for further storage. Thus, the shelf heights can be adapted to different goods and the space utilization can be optimized especially when unit loads of varying height are stored.

Shelving system In a shelving systems, each shelf has a continuous bottom (or grid) on which goods of any dimension can be stored. Owing to the flexibly adjustable height, different shelf arrangements and a variety of accessories the shelving system can be adapted optimally to the requirements of a manual order-picking where it is used as the standard rack type. The normal rack height is 2 m. When auxiliary means are used (e.g., ladders) also racks with a height of 3 m are used although the picking performance is reduced by additional movements. Furthermore, multi-level systems are used to better utilize the height of the room (mezzanine shelf). Here, the serving platforms are directly mounted at the rack supports (cf. Fig. 4.5).

Figure 4.4. Principle of a lift system [Photo: SSI SCHÄFER]

Honeycomb rack In a honeycomb system several racks are arranged directly one behind the other to create shelves with a special depth to take up long goods. As is usual with line racks a precise warehouse management and different storage and retrieval strategies can be realized by a definite assignment of shelves.

Depending on their use, the goods are stored directly in the shelf or on special load carriers (*long size trays* or *long size troughs*). Since these goods are long and sometimes heavy they have to be handled with special methods and facilities, e.g., special rack feeders, stackers or cranes. Because of the wide aisles the system is only suitable for large quantities of articles or goods.

Cantilever rack The goods are placed on cantilever arms which are mounted at vertical or inclined rack props. The cantilever arms can also be used to

Figure 4.5. Multi-level shelving system [Photo: BITO]

separate standing goods. Similar to pallet racks continuous shelves can be built for goods with different dimensions by inserting additional shelf plates. The cantilever system is ideal for the storage of long goods (pipes or rods) or boards. Usually these racks have a high loading capacity and can thus be used for universal storage purposes. Because of the flexible use of shelves, however, the stored goods have to be managed very carefully since the unit loads may be stored at any location. In line racks each article can be accessed directly.

The racks are served with different systems. In addition to the manual operation in case of light loads above all stackers and cranes are used. In some cases the cantilever arms or shelves may also be mobile to allow for a vertical access (from above).

Statical block rack

In statical block racks the unit loads are combined into a compact block. The rack arrangement allows for access from just one or two sides. Depending on whether the system is served from one or two sides the retrieval strategy is either LIFO or FIFO (cf. section 2.2.3). When units within the block have to be retrieved the goods have to be restored.

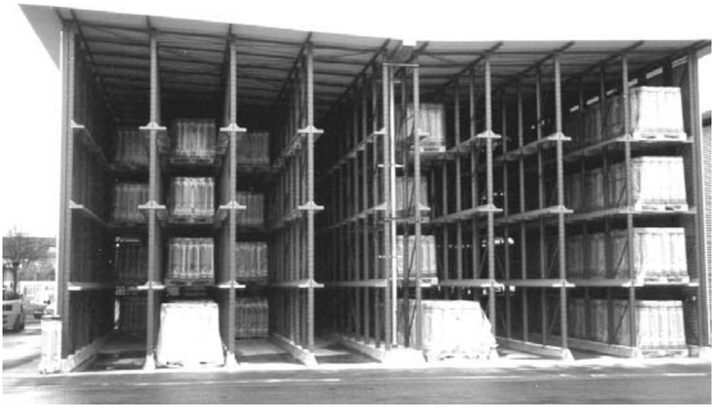


Figure 4.6. Drive-in rack [Photo: GEHRING]

The main advantage of this storage technology is that a very high space utilization without dynamic pressure can be achieved in line with a low space requirement (because of high storing heights). The unit loads are supported only at both fronts and therefore have to be of identical width. Because of the narrow aisles the unit loads have to be dimensionally consistent and stable. Owing to these characteristics this system is ideal for the storage of large quantities of a small assortment of articles.

Drive-in and drive-through racks The simplest kind of a statical block storage system is the drive-in or drive-through rack. The rack supports form vertical aisles for the rack feeders. Similar to the single bin storage angular profiles are mounted at the supports to hold the unit loads. When the unit loads are stored and retrieved they are moved slightly above the angle profile. The system is operated solely with front stackers and accessed from the side. Two kinds of movement are made to serve the racks:

Drive-in rack The unit loads are stored and retrieved at the same side. This inevitably requires a LIFO strategy.

Drive-through rack The units are stored and retrieved on opposite sides, this results in a FIFO strategy. The occupancy of the rack highly depends on the storage and retrieval frequency. The unit loads cannot be transferred in the aisles.

Operation and organization Because of these characteristics this system is not very suitable for the supply of different articles but for the storage of similar articles per aisle and the transshipment of complete units. The storage locations can be recorded and managed individually; this is difficult to realize due to the manual operation and the required identification of locations. However, this principle is ideal for the buffering of incoming or outgoing consignments.

Satellite racks This kind of rack is also called *channel rack with satellite*. Below the storage area of a channel there is a driving rail for a *satellite* or *channel truck* which can drive in this channel independently and has access to the first unit in each channel. Thus, more units can be handled than in a drive-in rack.

In most cases the drive-in method, i.e., LIFO, is used. But a channel can also be accessed from both sides. Different channel trucks are able to drive under stored units so that FIFO is theoretically possible. To achieve high occupancy with different articles the unit loads have to be restored frequently.

Operation and organization Automatic satellite vehicles are an ideal warehouse management tool. Since these vehicles serve the rack system independently and can be used in large numbers a compact storage can be achieved in line with a high throughput. However, the more diverse the mixture of articles in the channels (different unit loads in one aisle) the more single picks are necessary and the more restorage and organizational operations are necessary. This function has to be performed by the warehouse control. Furthermore, to improve the execution of retrieval orders it is necessary to move single unit loads near the retrieval point already in advance.

4.1.3 Dynamical racking system

For different reasons, warehouse systems are designed to be dynamical, i.e., the unit loads are moved between storage and retrieval. This offers the following advantages:

– Shorter order-picking distances and higher picking performance
– Higher transshipment in line with a compact storage
– Utilization of the benefits of a block and line storage

There are two basic principles for a dynamical storage: *Static unit loads in mobile racks* and *mobile unit loads in static racks*. The first group includes different kinds of flow racks.

Movable aisle systems

This type of rack is a statical line rack extended by a mobile platform, a so-called *subframe*. All kinds of rack systems like pallet, container, shelf or cantilever racks can be mounted on such subframes, but with a limited height (up to about 10 m). This allows for the sidewise moving of single racks so that a compact block is built and the aisles disappear (cf. Fig. 4.7). Alternatively, single racks can be moved (pulled) out at the front of the block.

Figure 4.7. Movable aisle systems [Photo: META]

In such a system each single unit load can be accessed when an aisle has been opened or the rack has been activated. The relatively slow motion of the large and heavy racks of 3-5 m/min limits the throughput. Therefore, the storage and retrieval strategy, which should focus on the minimization of rack movements, has a large effect on the throughput. When flow racks are used such a function should be part of the warehouse control.

Because of the above-mentioned features a flow rack system is ideal in cases where a high utilization of space is required in line with a low storage capacity, e.g., in archives or cold stores.

Rotary racks

Principally, rotary racks are conveyors (cycle or rotary conveyors) with an increased storage capacity. The aisles may be narrower than in a shelving system because the turning radius of the vehicles can be minimized. This method is often used in order-picking systems, which work according to the principle goods-to-man. The picking performance can be increased considerably by using several rotary racks at once.

Vertical rotary racks (paternoster racks) Turnable troughs are mounted horizontally at two vertically moving chains. On the one hand, they can take up small parts but also long goods (cf. Fig. 4.8). The picker is located at the central transfer point. But it is also possible to include several transfer points on different levels. By utilizing the full height of the room a relatively large number of articles with a small or medium number per article can be stored in a small space. The casing of the system serves as operator protection and prevents unallowed access to the goods.

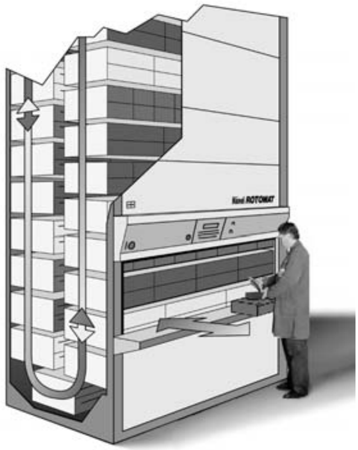


Figure 4.8. Vertical rotary rack (Paternoster rack) [Photo: HÄNEL]

The picking rate of a system largely depends on the building height and the access frequency. In order to achieve an acceptable performance the access sequence has to be adapted to the order of storage locations to avoid permanent reserve trips.

Horizontal rotary rack (Carousel rack) Single *storage fields* are mounted at a horizontally moving chain to take up a shelf column or similar. The picker usually retrieves the goods at the turning point at the frontend where several racks are arranged in parallel. To access a bin the chain is moved until the respective column or field reaches the retrieval point. Several racks (2-3) can also be mounted on top of each other or in very high systems (about 7 m), also vertically to save further space. The order-picking resembles that in a paternoster rack. Because of their length of up to 50 m carousel racks have a considerably larger storage capacity.

A special type of rotary racks are the so-called *rotary container racks*. These systems are mainly used for a short-term buffering, e.g., to consolidate partial orders in the shipping department. While in a classical rotary rack, single articles or units are retrieved in a rotary container rack the complete container is removed. For this reason, this type of system requires a high transshipping performance. Thanks to separate superimposed conveyor levels, each single level can be moved individually. During storage/retrieval the shelves to be emptied are placed above other shelves so that all operations can be performed in one step by a central exchange facility (step lift with multiple grippers).

Flow rack
By adding a conveyor level to a fixed block of racks, the unit loads can be moved within these blocks (also called *live storage*). For this purpose the unit loads are either transferred in a storage channel at the rear end of a rack, moved to the frontend where they are retrieved or are pushed into the channel at the front end where they are also retrieved (*push-in system*). In the channels the units are stored one behind the other and can be accessed only from the frontend of a rack.

According to the flow principle, storage and retrieval are separated and thus independent from each other. Furthermore, such systems inevitably work according to the FIFO principle (or LIFO in case of a push-in system) without additional controls. An advantage in addition to the compact storage is the possibility to store many articles with direct access and a guaranteed supply. The front side of the storage system may also be reduced what leads to shorter distances during order-picking and operation. Flow racks are used for light as well as for heavy piece goods.

Of operational importance is the fact that single items or single order unit loads are built in the single channels. It is almost impossible to restore units to get access to articles which are not stored at the front end of a rack. For this reason such systems are only suitable for order-picking where several units have to be stored without having to provide complete pallets.

Roller pallet systems These systems are a special kind of push-in system where the unit loads are placed on a roller pallet which are inserted into the rails in a storage channel. The unit loads within an aisle are coupled to trains by means of hooks at their frontend which is moved by moving the units at the aisle-side. Highly dynamical storage and transshipment systems and mixed channels can be built by providing sufficient storage capacity.

4.1.4 Pre-rack zone

Different tasks have to be performed in the pre-rack zones. Units to be stored have to be distributed to aisles, picking places in front of the rack have be be supplied with staging units and empty containers and, if necessary, the picking units have to be removed. In addition to this, retrieved units have to be transported to the shipping department or to be distributed to other aisles, during an inventory unit loads have to be retrieved for recording and restored afterwards, etc. This leads to several transport relations and cross-traffics. The performance of the pre-rack zone is influenced by two aspects: An efficient conveyor technology and a suitable control concept.

Used conveyors are roller conveyors, chain conveyors, turning deflectors or sidetracking skates. Depending on the rack feeder the goods are provided at both ends of the aisle or one upon the other (cf. Fig. 4.10).

Figure 4.9. Flow rack with roller conveyors for order-picking according to the pick-to-belt principle [Photo: BITO]

4.2 Transport systems

The term *transport system* generally describes technical material flow systems which are mainly used to move articles. For transport purposes outside the warehouse and distribution systems so-called *traffic means* are used which are not discussed here. This book focusses on transport systems for the transport of piece goods.

Transport systems can be divided into two groups depending on their operating method:

Conveyors work continuously and are in most cases stationary. They have a high transport performance measured in items/time units, and generate a continuous or quasi-continuous transport flow. They offer various possibilities for different routings and the permanent readiness to take up or transfer articles. Thanks to their continuous operation and easy use, they can easily be automated and the goods flows are easy to control.

Transporters are single devices which transport single or a few goods from a source to a target while traveling with the goods. This process is called *working cycle*. Depending on the construction type any point along a line, in an area or room can be approached. For this reason, these systems are ideal to serve a high number of transfer and pick-up points, to transport heavy goods and to cover large distances. Depending on the system design

Figure 4.10. Pre-rack zone with conveyor. To avoid a clashing of incoming and outgoing flows the conveyor routes are arranged separately and on different levels to save space. [Photo: KHT]

the requirements on the control and automation increase in line with the flexibility of use.

4.2.1 Conveyors

Roller conveyors Roller conveyors move the goods over stationary, turnable rollers (cf. Fig. 4.11). The dissipative motion resistance is reduced to the bearing friction in the roll and to the pressing of the goods while it passes the rolls. Thus, less energy is used for the transport of piece goods. Goods are moved manually, on inclined roller conveyors by gravity and by different motor drives. Heavy goods (about 1 t) are driven with chains, while light goods (<100 kg) are mainly driven with driving belts which are pressed to the bottom side of the rolls. In addition to steel rolls more and more plastic rolls are used.

To be transported by a roller conveyor the bottom of the goods has to be flat and sufficiently stable. The distance of the rolls has a considerable influence on the quality of the movement. The greater the distance the higher the vibration and thus the higher the strain on the goods.

When the roll shifts the goods this results in a slip according to the characteristics of the goods to be transported. For this reason, the positions of

Figure 4.11. Roller conveyor. In the inlet the rolls are tilted to adjust the goods as shown on the right side of the picture. [Photo: SIEMENS DEMATIC]

the goods on the conveyor line vary and cannot be determined. In automatic plants it has to be recorded at certain points by sensors (light switches).

When constructed accordingly these systems can also be used as buffers. Since the compression on the goods at the front increases when several goods are accumulated on a driven conveyor line the waiting rolls have to be deactivated when the accumulation has reached a certain length. Those types used for this purpose are called *accumulating roller conveyors*.

Owing to its simple construction roller conveyors often are a good alternative to simple transport tasks. But they are very noisy when the goods hit the rolls depending on the piece goods as well as the division and bearing of the rolls.

Belt conveyors The term belt conveyor is a general term for conveyors with a continuously driven element (belt, strap, lash or meshed wire) to take up and transport the goods. The goods solidly rest against the belt and do not move relative to the conveyor so that the movement is smooth, secure and gentle.

The most commonly used belt conveyor is the so-called conveying belt [83] with a belt wide enough to completely take up the goods to be conveyed. Rolls or plates are mounted below the belt to support the weight of the belt and of the goods. The goods to be conveyed should meet the following requirements:

Figure 4.12. Conveying belt: Helical curve [Photo: AXMANN]

– Take up tension forces with only little elongation
– Little bending resistance during turning
– Little friction on the lower side of the belt (running side) to reduce the required driving force; but sufficient friction to transfer the driving forces to the driving drum
– According to the use different friction values on the upper side of the belt (carrying side)

To meet all these requirements composite materials are used. Inlays from polyamide, polyester or aramide are used to take up the tension forces. Rubber or plastic insertions connect the outer layers which should also make the belt more resistant against wear and tear and environmental effects.

Conveying belts are used to transport light piece goods of up to about 50 kg. A suitable combination of belt and carrying construction can help to reduce the noise level drastically. In addition to the more common straight lines also curved systems can be realized by means of a lateral belt support (cf. Fig. 4.12).

Conveying belt systems can be controlled accurately and are highly dynamical because the transported goods seat solidly on the belt thanks to the high friction value. With this method, inclines can also be managed

Table 4.2. Types of chain conveyors for piece goods

Description	Function	Conv. speed.	Applications
Carrying chain conveyor	The goods lie on 1-3 parallel chains which run in special rails. The chains are deflected horizontally and vertically.	0.2 – 1.0 m/s	Horizontal and tilted transport; converter
Drag chain conveyor	The chain is a mere pulling device and sometimes runs below the conveyor level (*beneath floor level*). The goods are moved on rolls, rails, carts or fork lift trucks.	0.2 – 0.5 m/s	Transshipping centers, transport of heavy goods; manual transfer and pick-up of heavy units.
Jigger conveyor + paternoster	Load carriers to take up the goods are mounted between two vertical, at a paternoster parallel offset, rotary chains.	ca. 0.3 m/s	Vertical transport between floors.
Circular conveyor	Suspending load carriers are mounted at a chain randomly installed in the room which take up the goods. They are loaded manually or mechanically.	0.2 – 1.5 m/s	Floor-free transport of large quantities over long distances.
Power&Free (circular drag conveyor)	Similar to a circular conveyor but with an additional transport line for suspending load carriers which can be decoupled from the rotary chain.	0.2 – 0.5 m/s	Larger networks with different loops and buffer lines.

(cf.Fig. 4.12). Accumulating conveyors are built with very small belts or *lashes* where the goods are lifted by the lash during the accumulation.

Chain conveyor The characteristic element of a chain conveyor is a chain as carrying and/or pulling device. A *chain* has a high loading capacity, a small turning radius and can easily be equipped with additional elements. For this reason, it is widely used in conveying and storage systems (cf. section 4.1.3). The most commonly used chain conveyors are listed in Table 4.2. More detailed information about the single principles can be found in [40].

4.2.2 Transporters

While in a steady conveying system the complete line is designed for the heaviest or most sensible goods the dimensioning of a transporter system mainly focusses on the design of the transport means. This reduces investments in driving routes. Furthermore, different kinds of vehicles can be operated within

Figure 4.13. Chain conveyor: Flexible link conveyor for light piece goods (Carrying chain conveyor) [Photo: FLEXLINK]

a larger system as long as the requirements are met. Transporters are divided into two basic types according to their movability.

Freely movable The conveyors can freely move on the level. The third dimension is served by extensions or lifts. The free choice of motions puts high requirements on the vehicle control: Determination of the location in the room, operating and traffic situations adapted to the driving style or operation, anti-collision, etc. For this reason, the system is mainly controlled by the human workforce because their cognitive skills could only be simulated by a sophisticated control similar to that used in automated guided transport systems.

Controlled movable The conveyors are guided on (*floor-mounted*) rails which are propped up on supports or arranged *floor-free* outside the working area. Since an active steering is not required the control can be minimized. Most guided transporters are especially suited for automation.

Figure 4.14. Different floor conveyors in a warehouse [Photo: JUNGHEINRICH]

Floor conveyors
This is a generic term for all floor-bound transporter systems. The different tasks have a large influence on the design of the floor conveyors:

– Transport of piece goods over short or long distances
– Transshipment of piece goods between work equipment or exchange between work stations
– Stacking of piece goods and, if required, transport of stacked units
– Rack serving: storage/retrieval of piece goods in/from the rack and order-picking in the rack area.

Different types of floor conveyors are used depending on the frequency of one or several jobs. Simple transport jobs are performed by trucks or tractors with trailers (*transport devices*), which require an additional conveyor or a human being to be loaded and unloaded. *Universal devices* or universal transshipment devices can take up unit loads independently and move them to the transfer point or storage bin for automatic storage. Special *storage devices* are used to improve the space utilization and the transshipping or picking performance. These devices have narrow aisles, high-bay storage facilities or articles can be picked at the rack front. But at the same time they put high requirements on the quality of the ground and cannot be used universally.

Transport devices The mainly electrically driven trucks and tractors are used for regular transports over long distances. They are more economical

Figure 4.15. Tractor [Photo: STILL]

than fork lift trucks and bear fewer risks since the driver has a better view on his route.

For lower loads tricycles are used (up to about 8,5 kN tension forces, cf. Fig. 4.15) which have an undriven and controlled front wheel. Single devices are highly maneuvrable. In case of coupled trainsets (tractor with several trailers) the wheel guidance of the trailers is decisive for the width of the line. Trailers with only one controlled axle have a constricted cornering ability so that the width of the line has to be increased in line with the number of trailers. To keep the trainset on the path of the tractor the device has to be equipped with a four-wheel control with coupled axles.

Transshipment devices The main characteristic of a transshipment device is its ability to independently take up and transfer unit loads. Small transshipment volumes are handled with manual devices like carts, *trolleys* and pallet lift trucks. The widely used pallet lift trucks are designed for standard pallets (Euro and industrial pallet). When weights and transshipment frequencies increase, however, it is uneconomical to just use human resources so that the system is motorized in several steps.

In addition to the drive and lift the trucks are equipped with driver's cabins and seats to increase the transport speed. For order-picking the pallet lift trucks are also equipped with an elevating platform for the picker so that the second rack level can be served as well.

Stackers are used when access to higher levels or stacking is required in addition to the take-up and transfer of unit loads on ground level. The most common type is the fork lift truck (also front stacker, balanced fork lift trucks, cf. Fig. 4.16). A vertical lifting frame with forks is mounted on a three or four-wheel body, which takes up the load outside the wheel basis of the fork lift truck. To increase the loading capacity a counterbalance is fixed at the side opposite to the load.

Figure 4.16. Front lift truck loading/unloading a truck [Photo: JUNGHEINRICH]

To achieve the maximum lifting heights in line with low vehicle heights the lifting frames are telescopic. A U-shaped profile holds an internal frame which can take up further frames. The lift is moved by chain hoists which are activated by hydraulic cylinders at the side or in the center. The complete lifting frame is moved towards the goods and analogously towards the vehicle to transport the unit load securely on the forks. In addition to the admissible load and the maximum lifting height the free lift, i.e., the maximum lifting height of the forks without telescoping the mast, is a decisive factor for the functionality of a lifting mast since it determines whether the truck can also be operated in low rooms. Goods are always taken up and transferred at the frontend. The preferred driving direction is the forward direction (fork in front) so that in case of large loads the view may be obstructed. The large variety of accessories and extensions allows for a use other than the mere transport and transshipment.

Tricycles have a single controlled wheel at the side opposed to the load so that their turning radius is small. Four-wheel trucks have a swing axle on the side opposed to the load to increase the loading capacity. They are more stable (above all tilting resistant) but have a larger curve radius. Since the front fork lift truck has to turn by 90° to take up and transfer the goods the width of the warehouse aisles largely depends on the undercarriage and the possible load of a truck.

For in-house operations mostly electric stackers are used while diesel-driven stackers are used for outdoor activities. In mixed operation (in-house and outdoor) stackers driven with LPG (Liquified Petroleum Gasoline) can be used in addition to electric lift trucks. Another decisive factor for an outdoor use are the tyres.

Figure 4.17. Straddle stacker [Photo: STILL]

Storage devices this group includes all floor conveyors which are used mainly for rack feeding purposes.

Straddle stackers have two cantilever arms which are arranged below the load, e.g., below the forks (cf. Fig. 4.17). If the load has to be taken up from the ground, the distance of the wheel arms or straddles has to be adapted to the size of the unit load or the notches of a pallet. When the distance is too small the goods cannot be taken up from the ground and the unit load can only be taken up from a shelf when it has a lower free size of about 250 mm. Since the unit loads are handled or transported within the wheel base the fork lift truck needs no counterbalance and therefore is much shorter. The trucks normally have a three-wheel drive.

To take up units from the ground regardless of their width *fork lift reach trucks* have a mobile lifting mast which is moved along the wheel arms for load take-up (cf. Fig. 4.18). During transport the lifting mast is pulled back with the load to a stable position. Another alternative to straddle stackers are *reach fork trucks* which are seldom used today. The fork lift can be moved along the lift mast and extended to the front position for load take-up/transfer so that double-deep stored unit loads can be retrieved as well (cf. Fig. 4.19).

Figure 4.18. Fork lift reach truck [Photo: STILL]

The different types of wheel arm stackers require considerably smaller aisles than counterbalanced stackers. Since the load units are handled at the frontend the stacker has to be turned by 90° what calls for an aisle width of $\geq 3\,m$. To reduce the aisle width stackers are used which are able to take up the unit loads sidewise to avoid a turning in the rack system:

High-bay stacker These stackers are also called side loading fork lift trucks or three-side loading fork lift trucks. Telescopic forks or a rotating fork lift are mounted to the lift frame which can move the forks to both sides (*retractable swivel fork – cf. Fig. 4.20*). According to the type the retractable swivel fork can also be moved in the rack aisle.

High-bay stackers can serve racks up to a height of about 12 m. Unit loads can only be securely retrieved in such a height with the help of auxiliary means. For this purpose, the high-bay stackers are guided mechanically or inductively in the rack aisle. In front of the racks the forks are positioned by means of automatic shelf selection systems or camera systems.

Order-picking stacker Unlike a high-bay stacker the order-picking stacker or *high-lift stacker* is equipped with an operator's cabin which can be lifted by the fork directly at the lifting mast so that the picker can retrieve the unit loads directly from the shelves. The units are built by approaching

Figure 4.19. Reach fork truck [Photo: ATLET]

single shelves according to the man to goods principle. Some types are solely used for the picking of partial quantities and are unsuitable for the storage/retrieval of complete unit loads. For this reason, their fork is rigid and cannot be swiveled. Other order-picking stackers can be used to build partial units as well as to retrieve complete unit loads. These stackers have all the features of a high-bay stacker.

Side fork lift truck This type is ideal for the handling of long goods in cantilever arm racks. A crosswise traveling lift frame is mounted on a four-wheel trolley similar to a fork lift reach truck. These goods can be placed on the trolley for a secure transport. With this method also long goods can be handled quickly and longer distances can be covered without problems.

Four-directional lift truck Its construction is similar to that of a fork lift reach truck. But instead of the two rigid front rolls this stacker has driven rotary wheel motors so that it has at least two controllable wheel units (cf. Fig. 4.22). This allows the stacker to move in any direction, i.e., lengthwise and crosswise as well as to make traverse trips. The stacker must not be turned to take up goods since it enters the aisle crosswise to the load take-up position.

Figure 4.20. High-bay stacker with swivel fork lift [Photo: STILL]

Selection criteria In order to find the optimal floor transporter the complete task at hand has to be considered as well as the design of the warehouse and order-picking system and the transporter technology. Other aspects are the required transshipping and transport performance and their forecast development. On the one hand, the performance of a transporter system can principally be adapted to the number of transporters, but this rule does not apply for warehouses with narrow aisles where the vehicles cannot move or only one vehicle can be operated per aisle. Furthermore, the investment and operating costs increase in line with the specialization of the conveying system and have to be set off the achieved savings of space.

Because of the various side conditions, it is impossible to securely determine the profitability of single vehicles.

Operation and organization To efficiently use transporters pending orders have to be executed in such a way that prioritized orders are handled first, empty trips are minimized and existing resources are optimally utilized. A sufficiently experienced employee (stacker driver) may optimize his work sequence if he knows the order batches. And a dispatcher (control station assistant) can manually control certain optimization tasks.

Figure 4.21. Order picking stacker truck with liftable operator's cab [Photo: STILL]

This no longer makes sense when many vehicle units are operated in a dynamical environment with a short-term order lead time. In such cases the overall system can only be optimized by continuously recording the system status in line with the current job status of the single conveyors and their targets. To ensure a flexible and dynamical reaction each single unit has to be accessible at any time. For this purpose the system has to communicate with the decentralized operating units. Furthermore, the trips of the order-picking stacker between the retrieval positions in the rack system have to be reduced.

The optimization of the overall system requires a superordinate control system which controls the order management. The functioning and strategies of such a *transport management system* have already been discussed in section 2.3.3.

Another important factor is the documentation of finished work steps. Errors can be analyzed by means of reconstructed sub-processes so that arising problems (e.g., wrong article in the shelf, shortages despite of goods received, etc.) can be solved without an intensive error search. A prerequisite for this error analysis is the *tracking and tracing* of the material flow process (cf.

Figure 4.22. Four-directional lift truck [Photo: STILL]

section 2.3.4), which can best be achieved by considering the conveyors as temporary storage locations. Materials in transit are traceable at any time which is of special importance with regard to time-critical supply processes.

Automated guided vehicle systems Since these transporters are highly flexible they are difficult to automate. For this reason, such an automated system not only includes the vehicle control but also the stationary control with the vehicle and system control and the communication and safety facilities. The aim of such a system is to automatically transport the goods with transporters without the direct intervenience of the staff.

Regardless of the high technical expenditure automated guided vehicles (AGV) are used for different reasons:

- Transports in dangerous areas to improve operational safety
- Fulfilment of high requirements on the conveyors with regard to precision, repetition accuracy or strict observance of movement sequences, e.g., during the transport of large goods in restricted areas
- Reduction of labour costs, above all in multiple-shift operation
- Transparent processes
- Avoidance of transport damages
- Avoidance of floor installations in favour of rail-bound systems

Nowadays, all common transporters are also available as automated guided vehicles (AGV). The processes are optimized according to the same principles as for manual systems.

Cross-transporting vehicles
Cross-transporting vehicles, also called distribution cars or mobile carts, are floor-bound and rail-bound transport vehicles with a load suspension device (roller or chain conveyor) for the independent take-up and transfer of unit loads. They move between fixed transfer points, combine different inputs and outputs and serve as sorter and distribution device. They are frequently used for the distribution of goods in a pre-rack zone where they are an inexpensive alternative to conveyors.

Since cross-transporting vehicles usually are the connecting element between several facilities its performance has to be studied very carefully. To ensure a sufficient system performance the vehicles are optimized with regard to acceleration and speed. Another important aspect are driving strategies, which are adapted to the special application and give equal consideration to the choice of the shortest connections and to the date of the order. If required, an appropriate idle position has to be considered as well.

Rack feeders
Rack feeders are rail-bound conveyors for the serving of line racks (cf. Fig. 4.23). Thanks to rack feeders the constructional height of pallet racks and thus the space utilization could be increased. Typical systems reach a height of up to 50 m with an aisle length of 150 m.

Structure A rack feeder consists of two masts, a vertically traveling lift equipped with one or several load suspension devices (LSD) and a carriage. The rack feeder is guided by a ground and an overhead rail and driven by a wheel and disk gear or a synchronous belt drive, in some cases with ropes. Traction belts use stationary belts and mobile drives which are mounted on the carriage as well as stationary, external drive and mobile belts fixed at the stacker. To avoid vibrations the length of the traction belts is limited to the medium aisle length. Highly dynamical facilities may have an additional drive at the upper guide.

Rack feeders can be used for all types of racks served from the frontend (pallet and container racks as well as flow/ push racks). They can also be used in block storage systems, e.g., to transfer channel or satellite vehicles between the single channels.

Load take-up The type and number of the load suspension device on a lift truck have a decisive influence on the performance and functioning of the system as well as on the design of the system interfaces. The basic load suspension device principle is determined by the form and weight of the unit

Figure 4.23. Rack feeder for pallet handling [Photo: STÖCKLIN SIEMAG]

load. Pallets (loads >300 kg) are generally placed into the shelf with telescopic forks what avoids relative movements between shelf and unit load. The load transfer and positioning times are very long due to sequential movements (positioning of shelf – extending fork – lifting lift truck – retracting fork).

Because of the linear loading aid telescopic forks cannot be used for the handling of heavy loads which need a continuous and broad support (esp. air cargo – unit load devices ULD). For this purpose, roller conveyors are used as a load suspension device as well as at the shelf. This reduces the load transfer times but for cost reasons is only used for the above-mentioned unit loads without a sufficiently stable loading aid.

Light piece goods, however, (trays, containers; cf. section 4.1.2) can be transferred in many ways because they can among others, easily be pulled (a sufficiently smooth and level bottom of the unit load and a low-friction shelf storage provided). Different types of load suspension devices are shown in Fig. 4.24.

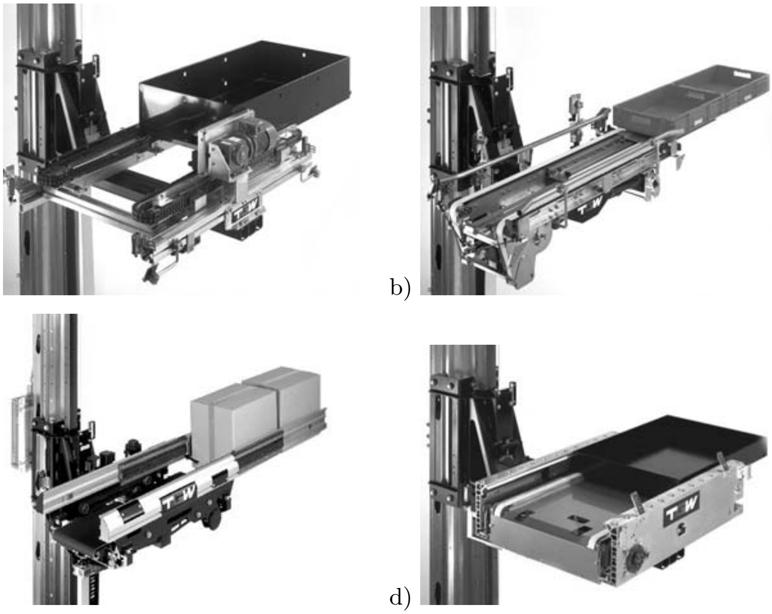
a) b) c) d)


Figure 4.24. load suspension device of a rack feeder: **a** Pulling device **b** Drive-in telescope **c** Gripper **d** Friction belt [Photos: TGW]

Control and operation The efficient operation of rack feeders requires the realtime control of their functions. For this purpose, own system controls in the form of autark material flow controls and subordinate facility controls are used to perform and optimize the movements on the field level. The strategies for the bin assignment (cf. Table 2.2) and retrieval (cf. Table 2.4) are represented in the material flow control as well as the department-specific bin management.

Special constructions Stacker cranes are the linking element between a rack feeder and a bridge crane. A vertical mast with the functions of a rack feeder mast (load guidance, lifting platform and load suspension device) are mounted at the trolley The main difference is that the system is suspended without ground installations so that it can also be used for movable aisle systems (cf. section 4.1.3). This principle resembles that of the so-called *Trans-Faster*, which is equipped with a suspended load suspension device mounted at the trolley. The trolley moves in an aisle on lateral rails like a channel vehicle so that several vehicles can move in an aisle on different levels.

The "lifting bar system" is a rack feeder rotated by 90°. The mast is a horizontal lifting bar guided on vertical rails. The lifting platform moves on this bar analogous to the trolley of the load suspension device. Owing to this technology, the number of carried drives can be reduced and the dynamic of the system increased although only in smaller warehouse systems.

Electric trolley conveyor

Electric trolley conveyors are floor-free automatic transporters. Electrically driven individual vehicles move along a rail which is suspended from the ceiling or mounted on pillars and powered via contact lines. The goods are transported by means of load suspension gears, gripping traverses or different kinds of hooks and lifting devices at the bottom of the vehicle. The load suspension gears may also be equipped with active load suspension devices (roller conveyors, chain conveyors) to optimize the load transfer and for the distribution of piece goods. In warehouse applications electric trolley conveyors are often used to connect different function areas (e.g., goods receipt and high-bay rack).

An electric trolley conveyor is a universal conveying system. It is loaded in idle condition so that also large and heavy goods can safely be transferred. With a driving speed of up to 2 m/s, it easily covers long distances and achieves high throughput rates. Extensive networks can be built by means of crossings and connections (turrets, two/three-way switches, parallel switches) as well as inclines/declines and lifting devices. In case of heavy loads and high throughput rates climbing aids can be used at inclines to reduce the technical expenditure at the single carriages which can manage 90° inclines with light loads. With a respective number of vehicles or parallel buffer lines also buffering tasks can be performed. The typical standard load for a vehicle is 300-1000 kg in addition to systems with a smaller or higher load. To increase the load single carriages are coupled by connecting at least one driven vehicle with one or several undriven carriages.

The necessary commands are transmitted either via the contact line or via radio or infrared transmission. In future, also contact-free transmission methods (inductive) will be used to minimize the wear and tear. In addition to the vehicle control (e.g., load transfer, distance control to preceding vehicles) a system computer controls the rail network (switches, etc.) and special sections to avoid deadlocks (right of way at complex sections) and to optimize the system performance (also block sections).

Special constructions Special types of an electric trolley conveyor are the electric carrying conveyor (ECC) and the electric pallet conveyor (EPC) (cf. [93]). Both run on floor-mounted rails similar to an electric trolley conveyor; the electric pallet conveyor moves on two parallel rails. The goods to be conveyed are transported above the carriage so that this system can better be integrated into certain applications. Both systems are controlled and organized like an electric trolley conveyor.

Cranes

Cranes are classical goods handling devices and mainly used to transfer goods between different transport means and for the handling of heavy goods. A major feature of all cranes is that they are floor-free and can move in nearly all directions without traffic routes.

Cranes can be divided into linear moving cranes with a cuboid working area and rotary cranes with a cylindrical working area. For outdoor use different kinds of *vehicle cranes* are used which are not discussed here any further [85]. Cranes with a cuboid working area are:

Bridge cranes Lateral rails are mounted above the working area (crane paths) which are directly supported by the building. The crane bridge spans the complete working area and moves on lateral front supports. The trolley moves above, parallel to and below the crane bridge and carries the load suspension device (rope or chain lift). Sidewise mounted *angle trolleys* induce a torsion moment into the crane bridge which is otherwise inflected. This minimizes the crane height and increases the working area. The movements and indication of forces into the crane path are achieved by steel wheels.

Suspension cranes Other than at bridge cranes the rails are mounted at the ceiling so that the crane hangs completely below the rails. This enlarges the working area since the trolley can also operate below the rail and can also move to other hall areas.

Stacker cranes Stacker cranes are based on a bridge or suspension crane but instead of a rope or chain lift they have a fixed, sometimes telescopable mast to prevent a swinging of the load suspension device. For this reason stacker cranes are highly suitable for rack feeding or dynamical diagonal drives.

Portal cranes If the crane bridge cannot be supported by the building it can be mounted on own supports on the floor. According on the size one support is a pendulum support to avoid tensions at the edges by heat or tolerances in the crane path. Due to the large mass and form of the supports fewer dynamical movements are possible compared to a bridge crane.

Wall crane The wall crane moves horizontally to a vertical wall and abuts into the working area. The rails are superimposed and designed only for small loads and goods of protruding length. The system can also be operated below another bridge crane.

Cranes with a cylindrical working area operate with a rotating cantilever arm. This may either be rotatable or equipped with a trolley to operate different parts of the working area. Indoor rotary cranes are only used in the form of slewing pillar cranes in industry and are of minor importance in warehouse and distribution systems. Outdoors different kinds of rotary cranes are used at harbor quays and construction sites.

Cranes in warehouse systems Cranes are mainly used in warehouses to handle heavy piece goods like steel coils, paper rolls, steel boards or special long goods. These goods are mostly stored in block storage systems which are operated from above. Automated systems are often used for this purpose because the operation and warehouse management is difficult for the reasons mentioned in section 4.1.1 (cf. p. 93). This requires an automatic load suspension device (magnetic or suction gripper or pliers) such as a suitable swing control. The crane must only be operated automatically in blocked areas without personnel.

In this case the storage bins have to be managed exactly since the goods have different dimensions and the bins vary according to the stacks. Furthermore, the sequence of stacks has to be considered as well. The throughput depends on the storage strategies and the required restorage. Cranes are also used for rack feeding (automatic as well as manual) whereas the requirements on the control are the same as on rack feeders. Furthermore, cranes are also used as transshipping devices in the goods receipt and issue where single goods as well as unit loads like containers are handled.

4.3 Sorting and distribution systems

Sorting and distribution system, in short *sorters*, are automatic devices which are used to distribute a large number of goods according to special criteria within a short period of time. Single processes can also be performed manually.

Because of increasing requirements on throughput times and operating costs but also due to changing shipping structures like frequent just-in-time deliveries of small quantities these systems are of growing importance. In the following the elementary structures and distribution principles will be described.

4.3.1 Applications

The requirements on the distribution process with regard to warehousing and distribution vary considerably depending on the application:

Transshipment for mail and parcel services Here, incoming streams have to be distributed immediately to outgoing tours. Regionally collected consignments are brought by vans to so-called *hubs* where they are distributed to other hubs as well as to regional deliveries leaving from the same hub. Because of the short delivery times the consignment has to be handled within a few hours. The priority of the sorting order depends on the distance to the destination.

The incoming and outgoing goods flows or assignments are mostly determined. The single consignments always have a clearly defined destination. The transshipment systems have to cope with high load peaks and pulsating delivery quantities. The range of articles is very large and inhomogeneous because they are manufactured by the shipper, i.e., outside the system. Guidelines refer to the weight and dimensions of the goods (admissible *circumference of the girth plus the length* often=length+2×width+2×height=3000 mm).

Order-picking systems The automatic distribution is an elementary part of two-step order-picking systems (cf. section 2.2.5). The customer orders which are consolidated into batches in the first step are separated for shipment. Since the number of daily customer orders often exceeds the possible number of final stations they are combined into batches which are executed successively. Thus, the articles are reassigned during each run. Furthermore, it is not possible to clearly assign a single unit since the same article may be requested for different orders[2] and may thus be assigned to different destinations. For this reason, the performance also depends on the application of suitable assignment rules.

Also one-step order-parallel order-picking systems can be rationalized with sorters. The consolidation of sub-orders into shipping orders (container sorting) puts other requirements on the distribution system: the goods to be sorted are consolidated quantities (containers, cardboard boxes), the throughput is relatively small and the quantities to be buffered at the final stations (often returns) are relatively large.

Sorting of returns Returns put special requirements on the goods receipt and have to be treated specially by the system (cf. section 2.2.1). When returned goods are restored they are distributed to the single warehouse areas. They can also be directly entered into the current order-picking process to avoid storage processes.

Baggage handling The most critical parameters at all airports is the minimum connecting time, MCT to guarantee for the secure transfer of luggage to connecting flights. At large airports these short process times can only be achieved with extensive conveyor systems with a sorting and distributing function. Special care has to be given to the goods to be sorted (baggage) which cannot be defined in detail with regard to their form and characteristics. These goods may be of any form and stability and have loops, belts, handles, etc.

[2] This condition is the intended aim of this principle to reduce picking distances in the first step (article-wise order-picking).

4.3.2 The basic structure of sorting systems

An evident characteristic of sorters is the distribution line with the corresponding techniques. A safe functioning requires the five up- and downstream features [91]:

– System input
– Preparation
– Identification
– Distribution
– System output

System input
The main task of the system input is to adapt the working characteristics of the sorter to the upstream system. Pulsating input flows, for example, have to be coordinated with the fixed distribution rate of constantly working distribution systems. This can only be achieved with lines for the unloading of incoming transporters offering sufficient buffering capacity.

In larger order-picking systems the just-in-time handling and provision of picking batches cannot be ensured without problems. It may thus be necessary to buffer several batches in front of the sorter to decouple the first from the second level.

Preparation
An automatic system generally is the more efficient the more limited the degree of freedom of the objects to be handled. This also applies to the distribution. Depending on the system sorting and distribution systems offer the following functions:

Decollation It needs defined intervals for the uninterrupted recording and allocation of a sequence of goods. This can be achieved, for example, by two series connected conveyors with a conveying speed of $v_1 < v_2$.

Adjustment To ensure the precise discharging the goods are adjusted on the distribution conveyor in different directions to the distribution loop. In case of bumpers or swivel roller conveyors the goods have to be positioned at the outer edge of the conveyor, e.g., by mounting roller conveyor sections with inclined or conical rolls. In case of single-slot conveyors the space in the trough and possibly at the terminal can better be used when the goods are placed lengthwise. For this purpose charging and feeding conveyors are arranged in a corresponding angle to each other or a manual feeder is used.

Another elementary prerequisite for the secure identification of the goods is a failure-free sorting. During different processes, however, the location of the identification mark may be random. To avoid the use of expensive all-side scanners (*scanner bridges*) the identification marks should best be positioned in direction of the scanner. This is mostly done manually but sometimes also by automatically tilting the goods into the correct position.

Applying additional information When piece goods cannot be differentiated by means of their dimensions, weight, design or barcodes they have to be tagged with additional physical or logical information.

Weight scan The weight scan is used to avoid an overload of the sorter or to identify picking errors.

Identification

The goods to be sorted are usually identified by means of barcodes, mainly in plain writing. These barcodes can be scanned at different positions, at the feeder or after they have been charged into the sorting loop. The earlier the barcode is scanned to better react to errors. Furthermore, the slow feeding speed improves the scan rate or allows for the use of several scanners.

If the goods are not marked with a code different manual identification methods can be used. Mail services, for example, manually enter the target code and then place the goods on the charger. In case of a mixed range of articles (with or without destination data on the label) more and more large systems operate with *telecoding*. Here, a camera is installed at the identification point which transfers the images to an external workstation. This not only reduces the error rate because of a better ergonomics but also increases the utilization of the human resources. The distance between the scanner and the decision point has to be large enough (processing time about 20 s.). *Voice-coding is of growing importance because it enables the picker to grip the goods with both hands.* Pick-to-voice systems are used in parcel distribution centers to enter zip codes or routing codes but also for the order-picking of articles (cf. section 2.2.5). Mail services more and more use the plain writing identification which integrates other methods (telecoding) in case of problems.

Distribution

The structure of a distribution system is categorized according to the following criteria:

– Arrangement in lines or loops
– Number and arrangement of the feeding stations
– Distribution and discharging principle

In line structures the goods are charged and distributed on a straight line (cf. Fig. 4.25, a+b). According to the distribution principle the conveyor may be deflected vertically at the end of the line which may thus be shortened. Undischarged goods (e.g., because of congested chutes) have to be returned to the charging station at the beginning of the line by additional standard conveyors where they have to be recharged.

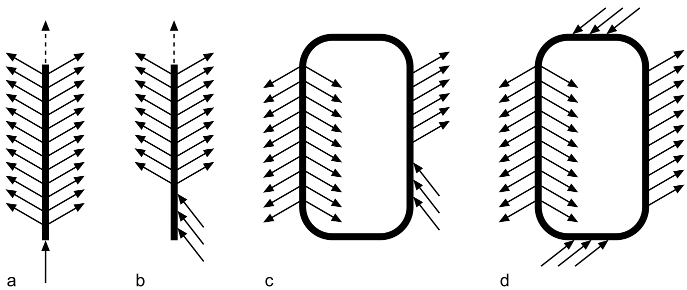


Figure 4.25. Arrangement in lines and loops

In a *looped system* the course is an endless horizontal loop (cf. Fig. 4.25, c+d). The terminals and above all the feeding terminals may be placed anywhere along the loop. If goods cannot be discharged at the defined terminal they can remain on the loop without renewed identification and charging and be discharged the next time they reach their terminal. Other than the line system this requires a long and expensive distribution route.

Goods can be charged from three different directions

1. Vertically from above
2. Horizontally in conveying direction at the beginning of the conveyor line
3. Horizontally from the side of the conveyor line

The goods cannot be charged freely but mainly according to the chosen distribution principle. Some distribution systems can only be charged from above, from the front or from the side (cf. section 4.3.3). A lateral and vertical feeding may lead to an increased system performance when several feeding stations are arranged in line. In addition to this, conveyor lines coming from other areas can directly be connected to the system. When a system is charged horizontally in conveying direction the complete flow has to be consolidated onto one conveyor line.

The increase in performance to be achieved by diagonally arranged charging terminals (cf. Fig. 4.25, d) depends on the pretreatment of the goods flow, the reaction of the single stations and the recirculation ratio (average share of cycling goods). In the ideal case the goods flow is presorted in such a way that the goods charged in a certain section can directly be discharged at the next chute without having to pass the next charging station. Thus, a multiple station occupance of the distribution line can be achieved. If, on the other hand, mixed flows are charged at different points of the loop the possible global increase in performance cannot be forecast because of the above-mentioned parameters.

On the one hand, a distribution principle can be described according to the assignment of sub-sections. In *single slot sorting systems* the distribution conveyor is divided into discrete sections (trays) which take up one piece good each. The theoretical distribution rate is determined by the division of the

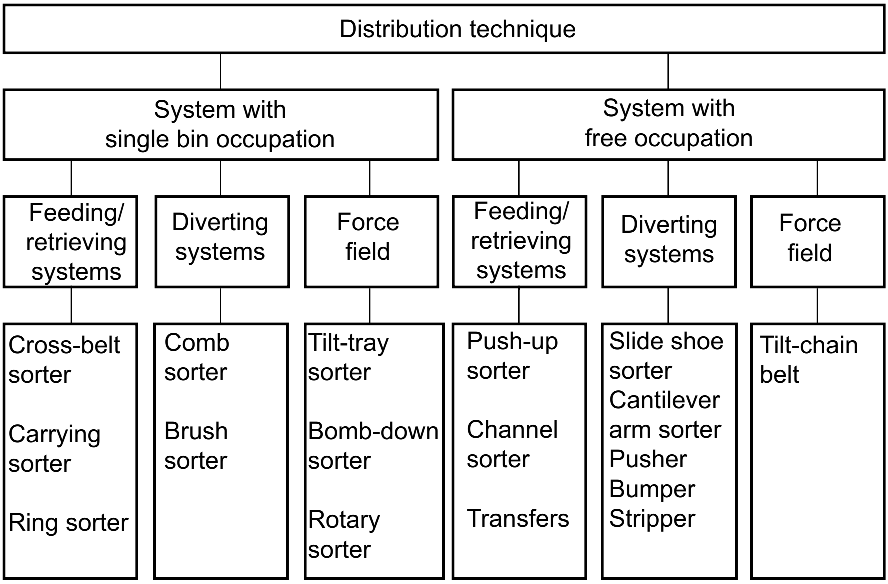


Figure 4.26. Systematics of the distribution technology

trays and the conveying speed. In certain facilities single elements can also be coupled to take up oversized goods. In facilities *with free occupancy* the distribution conveyor can be charged at random so that the distance between the goods can be optimized. This is of special advantage in case of largely varying articles.

On the other hand, the functioning of the distribution technology is defined by the active principle of the discharging mechanism. Current systems work according to the following principle:

Charging and discharging systems The goods to be sorted lie on a conveyor and are discharged by its movements. Since there is no movement relative to the conveyor the goods can be discharged precisely and gently.

Diverting systems The goods are lifted off the conveyor by a separate element. The positive fit between the good and the charger ensure the secure discharging. Depending on the design the strain on the goods may be considerable.

Force field-based systems The systems make use of the gravity and sometimes of the centrifugal force to discharge the goods so that the number of drives can be reduced. The discharging process depends on the characteristics of the goods (weight, density, situation of the center of gravity, sliding characteristics) and cannot be used for all types of goods.

The allocation of commonly used distribution methods according to this classification is shown in Fig. 4.26.

System output

There are various requirements on the *destinations* or *chutes*:

- The required order volume has to be securely stored
- The goods have to be conveyed safely after each breakdown at any degree of occupancy
- The goods have to be quickly discharged from the sorter to discharge as many goods as possible at the same destination
- The speed at the terminals and the resulting accumulating pressure must not be too high
- Little space required
- Easy discharging by the packing staff (ergonomic gripping height, no pushing of goods).

Because of their frequent use the system terminals require high investment costs. For this reason it is important to determine the most suitable technology and dimensions for a terminal. This is done by the following methods:

- Chutes (even chutes, stepped chutes, spiral chutes)
- Roller conveyors (driven or braked)
- Caster conveyors
- Belt conveyors
- Containers, troughs, roller containers
- Bags and shipping boxes

For cost reasons chutes are used for this purpose (cf. Fig. 4.27). To integrate the technology into the existing system, it should be connectable to succeeding processes. Active terminals such as roller conveyors or belt conveyors allow for an automatic handling. Passive terminals (often chutes, containers, etc.) are discharged by the staff and released for the following sorting job. This work step is of special importance for the performance of the overall system because the sequence of the terminals and routes between the terminals has to be optimized. In order to utilize the limited resource *packer* to the optimum optical status messages have to be integrated and, if necessary, suitable operating strategies have to be defined.

4.3.3 Distribution technology

Cross-belt sorters, tilt tray sorters and sliding sorters are frequently used for the distribution of classical piece goods. These sorters are described in the following. Other systems are described in [40]. In addition to the sorting of piece goods sorters are also used in mail services. Further special solutions are used, for example, for the sorting of hanging goods.

Figure 4.27. Terminals: chutes [Photo: BEUMER]

Cross-belt sorters Belt-conveyors are mounted on a chain crosswise to the conveying direction. These conveyors are designed for the maximum load unit size (cf. Fig. 4.28). Each belt conveyor can be moved separately. When the goods are charged sidewise the belt conveyor typically is moved synchronously to the feeding belt and the goods are transferred without a relative movement. The goods are discharged at both sides by activating the belt conveyor at the terminals. The belt conveyors are driven by positively driven controls as well as by single drives. In more expensive systems with single drives the driving and belt movements are completely decoupled so that not only the position of the goods on the conveyor can be readjusted but also the drop line as well as the position of the goods at the terminal. The complete chain is driven mechanically by chains, drag or worm gears or linear drives. The chains can be deflected horizontally as well as vertically so that the system can be used in loop or line systems.

Tilt tray sorters This single-slot system mainly makes use of the gravity force to discharge the goods. It also consists of a series of single elements (cars). Each car has a platform (tray) which can be tilted sidewise to the conveyor so that goods can be discharged at both sides (cf. Fig. 4.29). In case of the common sidewise charging the goods are decelerated by the friction in the tray (possibly supported by lateral guides). The system is activated at

Figure 4.28. Cross-belt sorter [Photo: AXMANN]

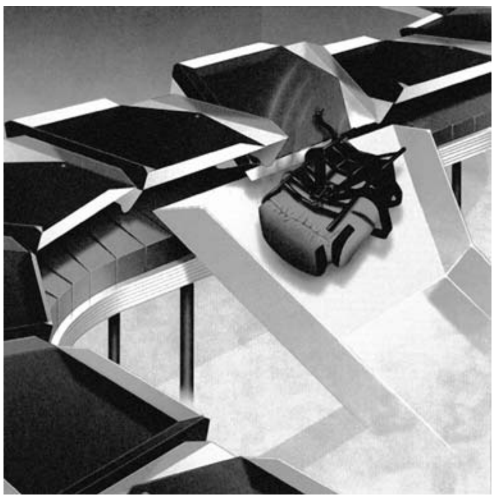


Figure 4.29. Tilt tray sorter [Photo: BEUMER]

the terminals by stationary switch cams. In addition to this, newer systems also have individual electrical drives for each tray so that the driving and discharging movement can be decoupled to achieve a more flexible goods transfer. The kind of the tilting movement can also influence the discharging process. The major aim is to reduce the width of the terminals. The drive and deflection are the same as in a cross-belt conveyor.

Sliding shoe sorter Two vertically rotating chains are equipped with bars or profiles (carrying elements) to take up and transport the goods. A movable element (slider) is mounted at each carrying element which is controlled by a negatively driven control below the carrying elements (cf. Fig. 4.30). The

Figure 4.30. Sliding shoe sorter (Posisorter (R)) [Photo: VANDERLANDE INDUSTRIES]

sliders can be moved on rails and switches in such a way that the goods on the carrying element are moved into the terminal. By preadjusting the sliders in the return run of the conveyor chain, the goods can be discharged at both sides. The positively driven discharge allows for the handling of a wide range of articles, except goods with straps and very small goods (<150 mm). The system is arranged in lines and the goods are charged at the front, horizontally at the beginning of the distribution line.

4.3.4 Control and strategies

A sorting and distribution system is a complex part of a material flow system and has to perform a variety of control and optimization tasks. An indispensable element is a separate control on the level of a material flow control for the control of type-independent functions.

In order-picking the system has to communicate cyclically with the superordinate control system to transmit the sorting criteria which change with each sorting batch. With this method the completeness of the sorting job as well as the stop criteria (e.g., because of missing goods and shortages (no-read) or faulty picks (no-need)) are checked. Finished and executed jobs have to be documented for the sake of a better transparency.

The efficient operation of a sorter is no trivial task. A variety of different control and process strategies are involved here, starting with the building of the order-picking batches over the stop and go of a batch, the prioritization of a destination assignment up to the control of the packaging system. Basic

Figure 4.31. Portal robots in order-picking [Photo: SWISSLOG]

interdependencies are described in [64]. A careful planning, e.g., with the help of a computer-aided simulation of the overall system, is indispensable.

4.4 Robots in warehouse systems

Robots are handling devices which can influence the orientation of a good in a room in addition to its situation. They have at least three axles and are used in warehouse systems for palletizing and order-picking.

4.4.1 Palletizing robots

Palletizing robots are, one the one hand, used to relieve man from monotonous and physically strenuous work which require intricate auxiliary means and thus long processing times. Compared to fully automated special facilities they are more flexible and cost-efficient for the handling of small work loads. Typical robots for these tasks are portal or articulated arm robots.

4.4.2 Order-picking robots

Robots are used in order-picking which are mounted on rack feeders and pick the goods from the shelf. Others operate in separate areas where they pick larger standardized units from pallet stacks. The technical expenditure and the performance depend considerably on the provision of pick-up units. The more precise this provision the easier can the robot be integrated into the overall system. These systems put large requirements on the positioning, decollation and handling of articles and thus demand time-consuming and costly processes.

5. Automation of the Material Flow

Automation comprises the independent operation of a technical system in line with high performance and economy.

With regard to computer-aided and thus mostly stand-alone functions, warehouse management is also a part of automation, however on a superior business management level (cf. Chapter 2 and 7). Here, above all the *strategic* and *anticipated* warehouse and distribution functions are automated. The automation of the material flow aims at controlling and supervising the *operative* handling of the material flow.

This chapter describes the basics of the material flow automation. Based on a hierarchical classification, we present first the main terms, requirements and tasks. The main elements of an automation technology are the *control*, *sensors* and *drives* to record and influence the material flow processes. The basic structure and functioning of the devices is described in the example of a typical application.

5.1 Basics of automation

Technological and economical aspects affect the technical design and operation of material flow systems and in many cases offer a possibility for automation. Although these requirements differ from case to case the goals are identical:

- Improved system performance (transshipment rate, shorter order lead times)
- Quality assurance (continuous quality of the products and processes, observance of deadlines)
- Cost savings
- Relief of personnel from uniform, strenuous activities

Whether or not these requirements are met mainly depends on the choice, dimensioning and arrangement of the function areas in the warehouse as well as the used conveyors and storage facilities (cf. Chapter 4). The main task of the automation technology is to ensure the smooth functioning of the single conveyors and storage facilities or their components and to coordinate interlinked systems.

Automation does not always make sense since manual solutions may be more simple or economical. An alternative to the in-house transport of pallets, for example, are the *manual* transport with a hand pallet truck, a *semi-automated* transport by stacker or the *fully automated* transport on driven conveyor belts and automated guided transport systems. Which solution should be preferred depends on the frame conditions set by the overall system and is not studied in detail here.

5.1.1 History of the material flow automation

Modern automation technology was largely affected by inventions in the field of electrical engineering and electronics some of which are listed below (cf. Table 5.1).

Table 5.1. Milestones in material flow automation

1947	Shockley, Bardeen and Brattain invent the transistor.
1958	Jack Kilby from Texas Instruments develops the first integrated circuit.
1967	The first card-controlled high-bay warehouse starts operations (first, still manually operated HBW 1962 at Bertelsmann company, Germany).
1969	Intel produces the first serialized micro-processor (Intel 4004), marketing of first programmable logic controller (AEG Modicon 084).
1970s	Electronically controlled automated guided transport systems are introduced in automotive industry. Use of programmable controls in high-bay warehouses, wide introduction of the barcode.
1981	IBM presents the first personal computer.
since 1985	Mobile electronic data memories are introduced in industry.
since 1990	Personal computers are more and more used in warehouse management for management and control.

Obviously, computer-aided data processing has had a considerable influence on the development of the modern automation technology, above all the development of microprocessors and the introduction of standardized computer architectures [24]. Almost all digital controls which are currently used in material flow technology are based on these principles.

Advanced processors and memories in line with an increasing integration of electronic components set the basis for systems the size of a check card and the performance of a common PC. In addition to central process computers, more and more mobile PC and handheld devices are being used. These are described in more detail in section 5.2.3.

5.1.2 Terms and definitions

According to the standard DIN 19233 automation is the independent performance of processes by means of suitable technical means [18]. Based on this process term the functioning of an automated system (cf. Fig. 5.1) shall be described in general.

A *process* is an activity within a system where material, energy or information is reformed, transported or stored.

Principally, the automation of a process requires *target values* which determine the desired procedure. This information can either be stored in the system or set by an operator independent of the situation.

The process and the respective functions of an automat often depend on the process status. For this reason, relevant information about the *process status* have to be recorded with *sensors* and *evaluated* in the form of signals.

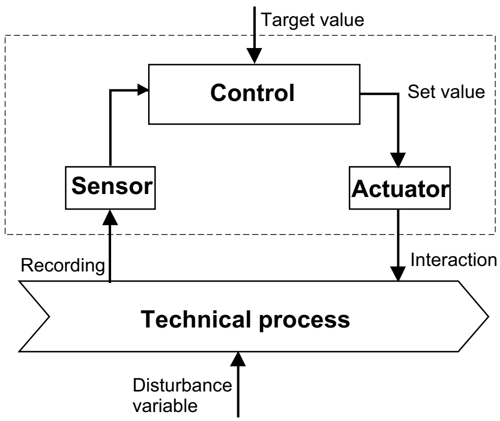


Figure 5.1. Active principle of an automaton

The input signals (state variables) are processed in a *control* according to defined rules (in a program) and at the end of the control returned to the process. *Actuators* then change these output signals (control variables) into process-specific parameters so that the process is influenced as desired. This is characteristic for all automated systems and can also be used for subordinate *sub-processes.*

One elementary material flow process is the translocation, i.e., the transport of goods. For example, containers provided in an order-picking area are filled with goods according to the order and then placed on a conveyor belt. This container is transported via switches, connections and dischargers either into an automatic container warehouse or onto a sorter which discharges the containers onto different lines in front of the goods receipt.

In a warehouse management system the transport of the container is considered as the superordinate process with the aim to discharge the container or to move it to a bin in the container warehouse. Sub-processes are the movement of a switch, a pusher or a rack feeder. The task of the control is to activate these components, i.e., to initiate the subordinate movement processes.

There are two different automation methods which describe the functioning of such systems [17]:

> *Open loop control* implies an open function where incoming target variables are inserted into the control line, i.e., into the control and the process. The control processes these target variables according to a stored logic and generated control variables for the process to be controlled. Disturbance variables which may affect the process can sometimes not be identified.
>
> *Closed loop control* describes a closed function where the status variables of a process are inserted directly at the beginning of the control and compared with the target variables to minimize, for example, disturbance variables.

In automated systems there is not always a clear difference between control and closed loop control since sensors always give a feedback to the control about the process status so that disturbances can be eliminated, if necessary.

A regulation in the narrower sense occurs in systems where processes are recorded continuously, for example the *warehouse regulation* at the axle drives of a robot (cf. section 5.4). To ensure the exact positioning of the unsteady axle load also under changing conditions while the robot follows the curves it is necessary to continuously check the axle positions and to compare them with the dynamical parameters (accelerations).

On the other hand, the automation of material flow processes with discrete piece good flows is discontinuous and the moved units are recorded only at predetermined points along the material flow. For this reason, the term *material flow control* is used as a synonym for material flow automation.

5.1.3 The structure of control systems

The technical design and structure of a material flow control is mainly determined by the type and number of the processes and sub-processes to be automated. It may either be *centralized* or *decentralized*.

In case of a centralized control system all control functions are performed by one control and all sensors and actuators of the operating means are directly connected to this control (e.g., process computer or SPS). Of advantage is the direct access to the overall system and the possibility to change the control process from a central point. When the control breaks down, however, the complete material flow system stops. In decentralized control systems the control functions are performed by different components with a limited scope which control a section of the material flow. These components exchange the control and process information of the different sections via communication systems.

A combination of both systems are *distributed* control systems where the decentralized components are coordinated by a superordinate system. Information is exchanged centrally via the superordinate system what requires additional communication.

In practice, distributed control systems with standardized communication networks and component interfaces prevail. An advantage compared to centralized control systems is the higher availability and investment security because owing to its modularity it has an open structure and is suitable for multi-client operation. When all single functions of the automated warehouse sections are controlled by one system, this may lead to complex overall systems. It has also to be considered that components and devices of different manufacturers often have to be combined into a functioning system. The possible variants and resulting requirements are legion. In order to design technically and economically suitable material flow controls which meet the requirements with regard to

– planning and projection,
– realization and implementation, and
– operation and maintenance,

it is necessary to structure the main requirements.

For this purpose the control functions must first be divided into process-related *single controls*, *section controls* and superordinate *master controls*. This inevitably results in hierarchical control levels with clearly defined tasks.

The possible hierarchies of automation systems are discussed in literature under different technical and organizational aspects [94, 39]. Logistic systems can be divided into the following levels according to the EDP systems or the fields of application:

– **Company management level** → Management of the company-wide logistics by production planning and control or enterprise resource planning
– **Factory management level** → Factory-wide coordination of all work fields, management of the operation of subordinate sections, levels of the warehouse management
– **Control level** → Division-wide control of facilities and work fields such as rack feeders, conveyors
– **Field level** → Recording and manipulation of processes
– **Material flow level** → Physical processes controlled by the superordinate level

A detailed hierarchical model was described by the transport forum of the German association VDMA[1] in its journal [94]. It determines the functions of the levels and interfaces to facilitate the planning and implementation of material flow systems and to coordinate the single manufacturers.

The hierarchical model shown in Fig. 5.2 consists of the master level, which is responsible for the order recording and dispatch as well as the inventory management and six other levels further down. This model represents the functions of material flow controls, but no fixed specifications since different levels can be combined depending on their complexity. Information are exchanged between the different levels from the dispatching WMS up to the performing elements on the field level.

Level 6: presentation and communication

This level represents the interface to the warehouse management. It receives transport requests transmitted by the WMS and sends them to the subordinate system control together with the system coordinates. It also sends feedbacks to order completion or error messages to the WMS. It is, furthermore, responsible for the operation of the plant, the visualization of the plant status and the logging of the processes. The most commonly used term for the computer systems used on this level is material flow control (MFC). Generally one or several interlinked industrial PC are used depending on the size of the plant.

Level 5: System control

On this level all transport processes within the material flow are controlled centrally. The warehouse management dispatches the transport requests as orders. Therefore, this level usually includes a model of the material flow system with all topological data, backlogs of orders and status information. It also includes transport strategies as basis for the optimized control (cf. Chapter 3) of transport orders in the system according to the plant status. This system control assigns the transport orders to the suitable conveyors and transfers them to the control of the respective subsystem.

Level 4: Subsystem control

The units on this level decentrally control all operations of the closed subsystems of a plant such as a high-bay warehouse or an automated guided transport system. Subsystem controls are also called subordinate controls and implemented in the form of an industrial PC or programmable controls (cf. section 5.2.2) of the upper class. For implementation and maintenance single sections can be semi-automated.

[1] Verband Deutscher Maschinen- und Anlagenbau – German federation of plant and machine constructors.

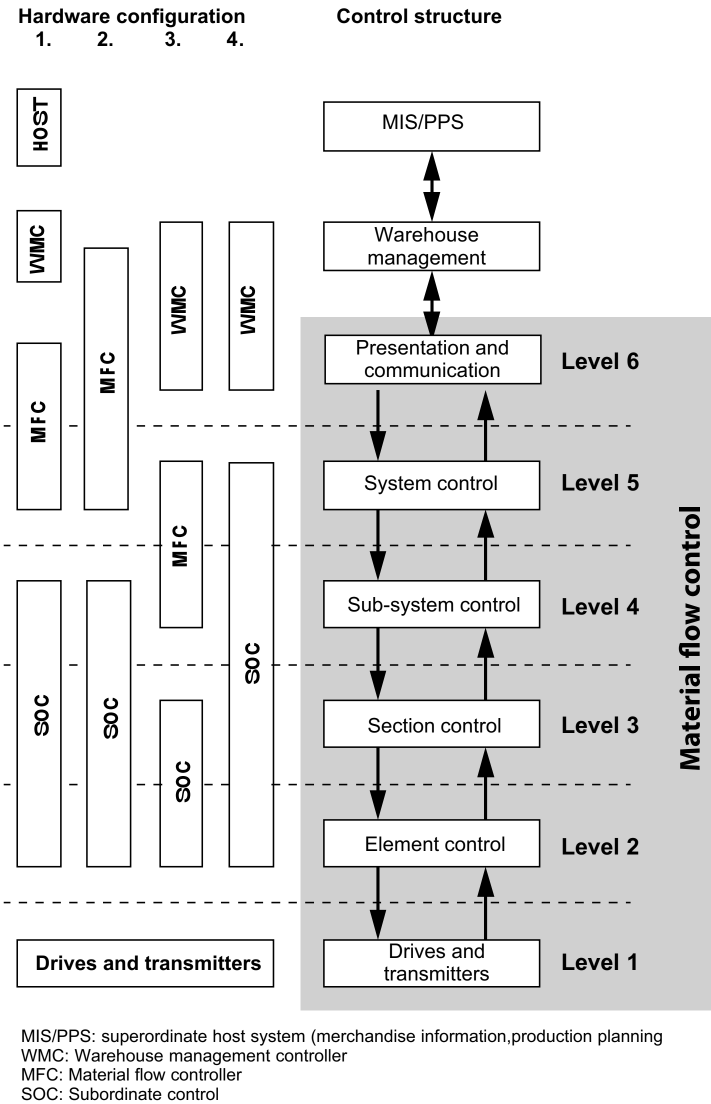


Figure 5.2. Hierarchical model for material flow controls according to VDMA 15276

Level 3: Section control

A subsystem control may be divided into different functional sections where independent section controls regulate the operation of the conveyors. These may be an automated guided transport vehicle or the section of a conveyor line. On this level the occupancy of the conveyors is managed, for example, by recording the containers on a conveyor belt which are identified by a scanner.

Level 2: Element control
The element controls control one or several drives of a conveyor element. They are either part of a section control or micro-controllers. This level is responsible for the safety mechanisms and the synchronization with neighbouring element controls.

Level 1: Drives and transmitters
Transmitters (sensors, cf. section 5.3) and drives (actuators) (cf. section 5.4) build the interfaces between the material flow control and the material flow level. Sensors transmit measuring values about the system status and the relevant processes to the control of the next superordinate level. The drives provide the mechanical energy for the transport movement.

In addition to the functional design, the planning (cf. Chapter 8) of a material flow control has also to consider and specify the interfaces between the levels or controls.

Above all the scope and time schedule of the information exchange between the control levels have to be determined. The requirements and the adequate communication means are studied in detail in section 5.5.

5.2 Control engineering

This section describes the basic functioning and structure of the control devices used in automation systems. Characteristic for these control devices are the presentation and processing of information as well as the *program realization*, i.e., the generation of necessary functions and active principles.

We want to point out that in addition to electrical or electronic controls also pneumatic or hydraulic controls are used for certain field-related control tasks which usually require just a few switches. But the technology of these devices is not described in detail in this book.

5.2.1 Classification of controls

To get an overview over, the multitude of different controls and the basic features which have to be considered during the project planning and choice are described below.

Presentation of information
Process signals which are recorded by sensors or transmitted to actuators and processed by a control are *analogous*, *binary* or *digital*.

Analogous signals are continuous in time. They are used, for example, for temperature measurements where the measured temperature is represented as a corresponding analogous voltage in a given voltage range, e.g., 0V up to 10V. Binary signals, on the other hand, can only represent two conditions (0 or 1), e.g., the supply of voltage. The information contents of a binary signal thus is one bit. Digital signals are built by a series of bits which represent a discrete value.

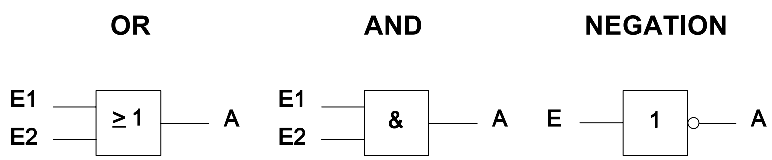
While in the past control information were represented by these different signals, e.g., in analogous computers, today only digital controls based on micro-processors are used. For this purpose, *A/D-converters*[2] are used which transform a recorded analogous voltage into time and value discrete digital values.

Types of functions

According to the German standard DIN 19237, the impact of functional and time-related aspects is characterized by two elementary types of controls:

In case of a *logic control* combinations of input signals are logically connected and represented without delay on the corresponding values of the output signals according to a if-then rule. A typical example for a logic control are, for example, safety circuits in plants. The drive of a conveyor can only be actuated when all emergency halts are released and no limit switch is activated. For this reason, the type of control is also called *interlock control*.

Logic controls can easily be modelled mathematically by means of the Boolean algebra as a combination of the elementary functions AND, OR, and NOT (cf. Fig. 5.3), which are also called *logical operations*.



E1	E2	A
0	0	0
1	0	1
0	1	1
1	1	1

E1	E2	A
0	0	0
1	0	0
0	1	0
1	1	1

E	A
0	1
1	0

Figure 5.3. Elementary logical relation (cf. [22])

In the simplest case the control is modelled in the form of a *truth-table* where all valid combinations of input conditions and the corresponding out-

[2] Short for analogous/digital converter

put signal images are listed. This results in an equation system which can be combined and reduced to an elementary logical equation and a *combinatorial circuit.* Extensive logic controls are modelled with automatic procedures which are described in more detail in [24].

The *flow control* describes the forced successive run of a program and thus of the process to be controlled. It consists of process and time-related steps.

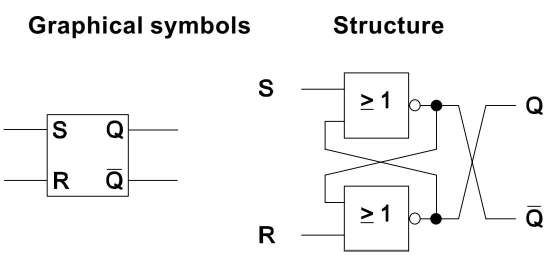
One example is the control of a pallet lift in a conveyor system. For the system control subordinate to the WMS (cf. section 5.1.3), the process consists of one single step, the movement between two levels under consideration of the source and target coordinates in the plant. In the control of the subsystem the process is staged:

> The pallet waits in front of the lift → Stopping and starting of the lift → Pallet is moved in → Lifting of the lift to the respective level and stop → Pallet is removed.

In this case, the control is divided into a logical sequence of single steps so that it is also called *sequential chain* or *sequential logic system.* The succeeding steps can start only when the preceding steps have been completed and all relaying requirements are met. Other than in combinatorial circuits in sequential logic systems the control process is either generated event-controlled by external relaying conditions or by a pulse generator. Suitable control elements for this purpose are:

– Latches (flip-flops)
– Timing elements
– Counters (shift registers)

Latches have to be able to store two signal stati (logical "0" or "1") and thus are bi-stable elements. A basic form of this kind of memory is the so-called *RS flip-flop* which has two inputs and two outputs (cf. Fig. 5.4).



Truth-table

S	R	Q	$\bar{Q}$
0	1	1	0
1	1	X	X
1	0	0	1
0	0	X	X

Figure 5.4. Structure of an RS flip-flop

Flip-flops consist of two elementary logic operations, in this case NOR gates[3]. The RS flip-flop differentiates three possible conditions according to the truth-table:

Setting: At “1”-level and S-input the output Q is set at “1”-level.

Storage: If the S-input then again has a “0”-level the output Q remains unchanged.

Resetting: If the R-input is connected with “1”-level the output Q is set to “0”-level.

If both inputs have a “0”-level the condition of the flip-flop is undetermined.

Programming of controls

In practice controls use a combination of both functions. Thus, the interlocks within a sequential chain often are a combination of stored system conditions and process-specific signals.

The required logical and process-related functions are called a *program*. According to the German standard DIN 44300, a program is an instruction required for the performance of a task with all necessary agreements. But the use of programs does not necessarily result in a software solution. The term can also be used for discrete, i.e., hardware-based controls.

Wired program systems (WPS) are wired or fixed programmed controls for defined automation tasks which usually cannot be changed or adapted afterwards or only at great expense. For this reason these controls are mainly used in small systems with only few inputs and outputs and safety functions (emergency stop, limit switch). They consist, for example, of relays and contactors and can thus only be used for the processing of binary signals.

Programmable logic controllers (PLC) contain the control functions in the form of program codes which are either freely programmable (software) or programmable with interchangeable memory (firmware, EPROM modules[4] so that they can easily be changed later on. In addition to the automation devices of the same name, the group of PLCs also includes controls based on industrial PC and micro-controllers.

This term is not only in its original sense for the group of controls but also for automation devices since these are widely used. PLCs are mainly used on the lower and medium control levels.

5.2.2 Programmable logic controllers

In order to avoid confusion, in this section the term PLC is used as synonym for the automation device. The standard DIN EN 61131 (IEC 1131) defines a programmable logic controller as [37]:

[3] They may also consist of NAND gates; in this case the signal level inverted at the input and outputs has to be considered

[4] Erasable and programmable read only memory, e.g., erasable by UV light.

> "A digitally operating, electronic system to be used in industrial environments with a programmable memory for the internal storage of user-specific control instructions for the implementation of specific functions such as the logic control, flow control, time, counting and arithmetic functions, in order to control different kinds of machines and processes by means of digital or analogous input and output signals."

PCLs have been the long-used core of industrial automation systems because of their reliability and robustness. Owing to their modular structure and large variety of possible configurations, PLCs can be used for various applications. Furthermore, they are inexpensive and easy to operate.

PLCs are mainly used to automate flow-controlled processes thus the name *automation device*. Modern programmable logic controllers are advanced logic controls which originally were designed as connector controls. In 1968 a group of engineers at General Motors developed the first PLC; since 1979 the controllers of the Siemens series SIMATIC S5, and of the succeeding series S7, have been used as a worldwide standard in many industrial areas.

System structure
The basic PLC consists of the following function groups:

Power supply: Provides the power for the device and protects it against power surges.
Communication: Acts as user interface and allows the operator and the programmer to program and maintain the PCL.
Interfaces: The PLC is connected with the plant section to be controlled or the processes and the superordinate systems via input/output elements and interface processes. At these interfaces sensor values are read, actuators are activated and process conditions are transmitted to the master computer or managing tasks are accepted.
Signal processing: The core of a PLC is the signal processing in the CPU[5] where the input signals are processed, the logic control connections are computed in a determinable time, i.e., *realtime* and transmitted to the output components.
Storage: Flow control functions are directly connected to the *storage functions* where communication and interface data are evaluated and new values are computed.

Structure
As already said, a PLC is a modular system allowing for the easy exchange of certain modules or the integrated functions. Generally, all modules are

[5] *central processing unit*

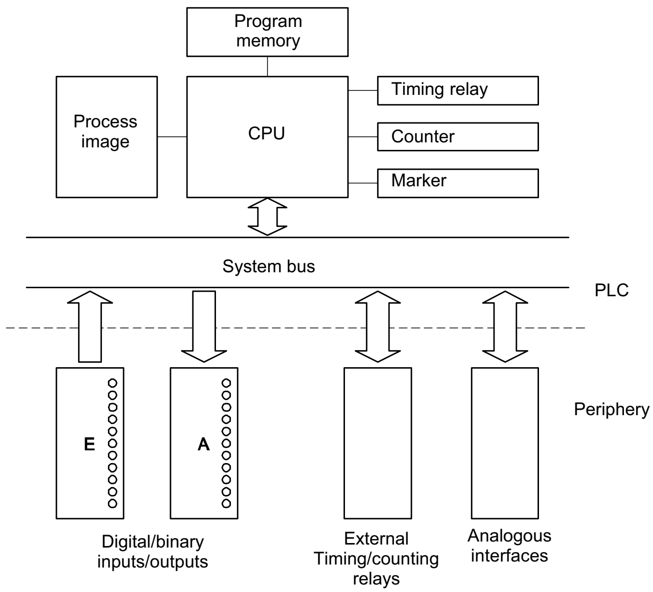


Figure 5.5. Components of a PLC

mounted on a carrying frame which is responsible for the connection of the elements as well as for the power supply via the system bus installed at the rear panel.

The *processor assembly* consists of one or several processors, internal registers (flags) and a memory for the program and the so-called *process image*. In addition, it also serves as connector for external devices such as programming, operating and output devices.

In the *input assemblies* binary signals are processed by generators for further processing. Each input is electrically isolated from the internal system by an opto-connector and equipped with a delay module for a secure signal recognition and to avoid false releases because of “shattering” buttons or switches.

Output assemblies emit the binary signals computed by the central unit. Here as well, each output is electrically isolated from the system by an opto-connector. Circuit-breaker switches are mostly integrated and higher voltages can be achieved by coupled relays or power switches (Thryristors, Triacs, etc.).

Special assemblies include analogous input/output modules. Here, incoming analog values are converted by A/D converters for the internal digital processing. On the other hand, digital values by converted by D/A converters into the required analog voltage.

Further assemblies are counter modules for quick events, display modules, computer coupling modules, diagnose modules as well as bus couplers for different *field bus systems.*

Functionality and programming
All programmable logic controllers operate *cyclically* with stored programs. At the beginning of a cycle a *process image* is generated from the input signals and stored in the PLC memory. This image is processed by the program and transmitted to the outputs. The process image is updated once before each program cycle and remains unchanged throughout the run to avoid inconsistencies due to changed input signals.

This method is very fast but may lead to different runtimes per cycle since the software may also include branches. The PLC software can be divided into different components with three hierarchical levels according to standard IEC 1131-3:

– PROG: The main program of the PLC. Here, the periphery and the memory (global variables) are managed.
– FB: Function blocks access input and output variables and manage the *tags* (memory elements).
– FUN: Functions (also calls sub-components SC) are sub-programs and also control the input and output variables but with a memory function.

A component can only be accessed hierarchically from the superordinate component.

The programs are generated with the text and graphic tools of the PLC providers which are listed in the standards DIN EN 61131 or IEC 1131, part 3 [37]:

The *instruction set* (IS) resembles the internal computer-code of the PLC the most. The program is a list of single instructions (cf. Fig. 5.6) which are executed in the given order during the program run. An instruction consists of the logic operation and the respective operator. Owing to the machine-oriented programming with IS, a high processing speed can be achieved; however, in case of extensive control tasks the program-code may become very complex.

The programmer may find it much easier to program the control in the form of a *structured text* (ST) which allows for abstracted and intuitively understandable commands. But at first the source text has to be converted by a compiler (cf. Chapter 7). As a result the machine-code is relatively unaffected, which may lead to a less efficient runtime. An advantage, however, is the compact and clear formulation which can also be used for more complex tasks. Furthermore, the standard IEC 1131 also provides for the use of compilers and thus for the programming in high-level languages such as C.

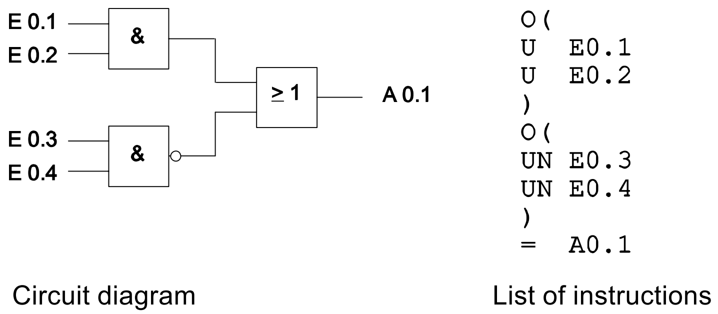


Figure 5.6. Example of the programming of logic operations with IS

In addition for text entries also graphical entries can be used for programming purposes. The *ladder diagram* represents all functions in the form of circuit symbols similar to the presentation in circuit diagrams. The relay wiring was adapted from electronic engineering so that a program can be represented like a circuit diagram (cf. Fig. 5.7). The symbols *open contact* and *closed contact* are used to mark the *consumers* (such as drives, displays).

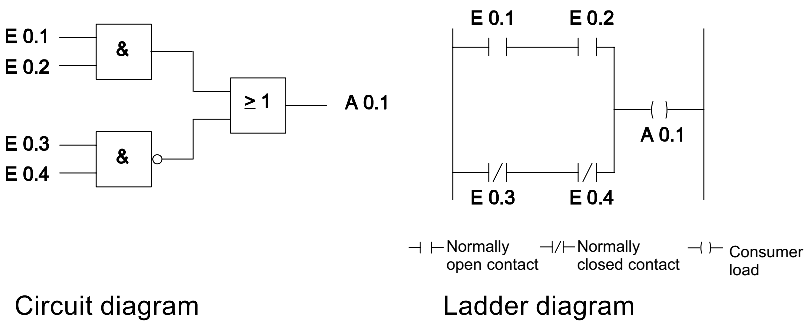


Figure 5.7. Example of the programming with a ladder diagram

The *continuous function chart* (CFC) or the *function block language* (FBL) can be used to process integers or floating point values. The circuit symbols used in the ladder diagram have been replaced by logic symbols of the Boolean algebra.

The *flow language* (FL) can either be used in graphic as well as textual form and is even more complex than the languages described earlier. It can be used to divide a complex program into simple blocks or units and thus to get a clearer structure.

5.2.3 Computer control

The term *process computer* includes all types of microprocessor-based systems which are used in automation for the control and process management. In this context the PLC (cf. section 5.2.2) is a kind of process computer. There is quite a variety of other computer systems which are used in different partly very special control areas.

Process computers generally are all kinds of programmable digital computer which offer the following functions::

– Realtime[6] data recording, processing and output
– Input and output of process signals in the form of electrical signals
– Single bit processing

The architecture of process computers is similar and consists of one or several processors, memory modules, data and address bus as well as I/O modules to couple the periphery. The different designs are classified into

– Single-chip computers (micro-controller, embedded systems)
– Single-board micro-computers
– Industrial PCs, plug-in computers

Single-chip computers are integrated micro-controllers that include all components (processor core, program and working memory, I/O modules as well as A/D converter, counting and timing relays) in one chip case which is equipped with connections for the power supply, signal lines and, if necessary, a serial interface to the *in-Circuit programming*. Because of its size the capacity of the program and working memory is limited to one kilobytes up to some 100 kilobytes. Since micro-controllers usually have a small instruction repertoire, such chips are mainly programmed in assembler code (cf. Chapter 7) or by means of specific high-level language compilers. Since the control functionality is in any case processor-specific the programmer has to find a suitable program for time-critical processes which are determined by the cycle time of the process commands to be executed. Micro-controllers are used, for example, for machine-oriented control functions in sensors or actuators or as communication module in field bus controls.

Single-board computers are based on a micro-controller as well as discretely on single integrated circuits. In addition to actual computer components the board[7] holds the signal periphery for the level adjustment and interfaces for the connection as plug-in computer or screwed contacts for a free wiring.

[6] German standard DIN 44300 defines realtime capability as the "operation of a computer system with continuously available programs for the processing of data so that the results are available within a given period of time."

[7] Conductor board as base for electronic components or modules which are connected via electrical conductors.

Table 5.2. Characteristics to be used for a sensor classification

Characteristic	Specification (categories)
Type of measurement	Presence, length, angle, speed, revolution, identify, ...
Dimension	0-, 1-, 2- or 3-dimensional
Recording of measurements	Tactile, not tactile
Measuring principle	Direct, indirect
Measuring method	Optical, mechanical, magnetic, inductive, ...
Presentation of measured values	Analogous, binary, digital

Industrial PC are similar to conventional personal computers; however, the casing and components are designed to meet the special requirements of the operational environment and availability. The signal periphery is connected with the PC by plug-in cards which allow for interface connections similar to a PLC.

Industrial PCs are used for the simple dispatching in order management and process supervision in subordinate controls up to the direct control of processes by means of sensors and actuators. While the first two application do not need deterministic process response times of a few milliseconds and thus put no requirements on the time behaviour of the installed operating system a *realtime capable* operating system is indispensable in process-oriented controls (cf. Chapter 7).

5.3 Sensors

To be able to influence an automatic system and its processes it is necessary to understand the relevant system states. The status of a system is described by a number of physical parameter which have to be identified by sensors and changed into a suitable (electrical) output value which can be processed by the control.

5.3.1 Sensor classifications

Sensors are classified according to different characteristics like the physical measuring volume, the measuring principle or the representation of measured values at the sensor output. This classification is also used as basis for the choice of a device for an automation system.

The tasks of a sensor are defined by the physical measurements to be recorded. In this context, the dimension represents the scanning range of the sensor or the presentation of the measured values. The classification into tactile and non-tactile sensors shows which kind of measuring device has direct mechanical contact with the process and thus is prone to wear and tear. In the following, the typical types of material flow sensors are described, classified according to the physical measuring principle under consideration of technical designs and interfaces.

5.3.2 Mechanically operated sensors

Switches and push-buttons are often used to supply low voltages. This sensor group is characterized by a simple and maintenance-free structure and is used for many industrial applications. An external force activates an electrical contact in the sensor which turns on the power supply at the sensor connection.

Mechanical sensors designed as break contacts, open contacts or toggle switches supervise these functions so that cable breaks or sensor breakdowns can be detected by a potential drop in the signal chain. Due to their construction and the tactile operation switches and push-buttons are prone to mechanical wear and tear but achieve a relatively high frequency (about 500,000 - 10^7 switching cycles).

The contacts generally allow only for small voltages up to some 100W. Typical is a connection at the input module of a PLC. Here, an auxiliary supply provided by the control and connected by the sensor is evaluated by the PLC as a binary signal.

Push-buttons and switches are used to detect the final positions of linear drives or load take-ups at rack feeders (limit switch) as well as as safety element on vehicle bumpers. Furthermore, they are part of control panels and safety circuits (emergency stop).

5.3.3 Optical sensors

Optical sensors are widely used in material flow systems for measuring tasks. Their range of application reaches from a simple sensor for attendance checks (light barrier) up to the automatic check of the position and environment conditions of autonomous vehicles.

Optical sensors have a somewhat similar structure and consist of a light source (sender) and an optical detector (receiver) which converts the intensity of a light beam which is reflected or broken by an object into an electrical signal. Specifical differences are the arrangement of the optical path and the evaluation in the receiver.

Because of their measuring principle optical sensors work contact-free and thus free of wear. In practice, problems may occur due to soiling and condensation at the optical parts as well as external light.

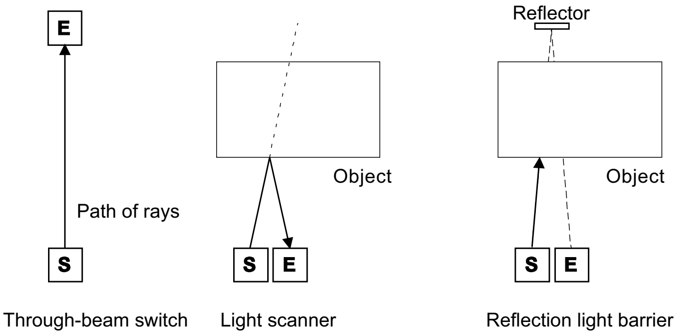


Figure 5.8. Optical recording with light barriers and scanners

Attendance check

For the automatic recording and counting light barriers and light switches are used (cf. Fig. 5.8) which generate a binary output signal. The sender consists of a light source (LED, visible and IR light) and a focussing lens, which bundles the beam. The receiver includes a light-sensitive element (phototransistor) and is connected to a sensor, which switches when the intensity of an incoming light beam exceeds or falls below an adjustable threshold value. This threshold value can usually be set by a potentiometer to adjust the arrangement of the application environment.

One-way light barrier: Here, the sender and the receiver are separated and mounted, for example, at opposite sides of a conveyor belt to record and count passing objects. For this reason, one-way light barriers are also called *through-beam switches.*

When the light beam is disconnected a switching operation is started in the receiver. The distance between the sender (S) and the receiver (R) can be set at random, but it has to be ensured that the optical beam is exactly adjusted. Because they work over large distances one-way light barriers are also used to secure gateways and entries to machines. Here, the sender as well as the receiver have to be supplied with an auxiliary voltage.

Light scanners: In these devices the sender and the receiver are mounted in a casing. To start a switching operation, the light beam has to be reflected by a passing object and guided to the receiver. When the beam hits the receiver, the switching operation is started. To avoid faulty switches by reflected external light infrared diodes and transistors as well as filters are used.

The function of the light scanner depends on the reflecting characteristics of the objects to be scanned. The light has to be reflected off an object surface so that the intensity of the beam in the receiver is high enough also when the object is tilted. Highly light-absorbing or reflecting and very smooth surfaces (foils) may limit the usable scan (scanning range) considerably or even make a measuring impossible.

The scanning range of light scanners is usually limited to a distance of $\leq$ 2m.

Reflecting light barriers: These are structured like light scanners but the light beam is not reflected by an object but by a reflector on the opposite side, which permanently returns the beam to the receiver. When the beam is disconnected a switching operation is started. The principle is similar to that of a one-way light barrier. As with light scanners, the reflecting surfaces of passing objects are problematic but for another reason: The beam path — and thus the object detection — may be impeded by reflections at the object.

Light grids: Single light barriers are arranged in lines in cascades. The sensors of this system bundle the single binary output signals into one digital signal. Light grids are used to control the height of packages and pallets and to secure the warehouse aisles.

Path and angle measuring

The through-beam or reflection principle is used in a modified form to determine the position of the axles of conveyors and drives.

The positioning requires an optical scale on a driving axle or on the path which can be scanned by the sensor either *incrementally* or *absolutely*.
In case of an incremental scan a single sensor element (light barrier or scanner) scans a path of alternately reflecting and light-absorbing fields (reflected light method) on the scale or fields with gaps (through-beam method) of the same width. The relative path or angle and, based on the measuring time, also the speed can be determined by counting the phase changes at a fixed reference point.

It has to be ensured that in incremental systems the movement of the drive is always recorded by a reference tag so that in case of a blackout the information will not be lost.

Absolute scales can have one or several tracks. In contrast to incremental systems the absolutely coded measuring systems have to be referenced just once since a clear binary coded position value referring to a virtual zero point can be read at any point on the scale.

Single-track scales have light-dark fields or gaps of different widths arranged in the measuring direction. The fields represent a non-iterative sequential code [69], which is scanned by a sensor line consisting of sidewise

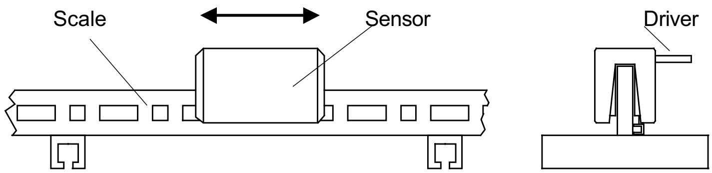


Figure 5.9. Optical absolute coded path measuring system

arranged light barriers and transmitted at the sensor output in the form of a digital value.

The measuring system is used by a telpher line or rack feeders as positioning scale, for example with a steel rail mounted at the driving path (cf. Abb. 5.9) The signal processor integrated in the sensor head continuously determines the absolute position of the sensor according to the scanned bit pattern and transmits the resulting value via the integrated serial interface. The resolution is less than 1 mm at a maximum measuring distance of several hundred meters.

In multi-track systems each track of a bit corresponds to a code consisting of n tracks spaced one above the other and thus builds 2^n discrete clear position variables. This principle is used in absolute coded angle transmitters which are flanged to the axle of the positioning drive and detect the position of the motor shaft. The position variables are also transmitted serially.

Measurement of lengths

Transit time techniques are used for the optical measurement of lengths. With this method, the time between the generation of a light beam in the sender and the reception is measured. Sender and receiver are mounted in a casing and together with the measurement processor they form the sensor which determines the distance to the objects. For physical reasons this method requires a coherent, i.e., unbroken laser beam generated by a laser diode. Industrial sensors achieve a measuring accuracy of some millimeters and modern high-resolution systems even in the range of micro-meters.

These systems are used alternatively to scale systems in high-bay warehouses for the positioning of rack feeders. In these cases, the sensor is mounted on the carriage of the rack feeder and directed at a fixed reflector at the end of the aisle. Because of this arrangement the laser distance measuring devices are limited to straight moving facilities.

Laser scanners are also based on this measuring principle, but the laser beam is additionally reflected fan-shaped by a rotating tilted mirror above the casing and generates a planar scan field (cf. Fig. 5.10). The sensor scans the environment in preset steps and at each cycle generates a set of single measuring values corresponding to the distances in the polar coordinates. With a local resolution of 0° - 1° and an average environmental distance of 10 m objects with a minimum width of 10 cm can still be detected.

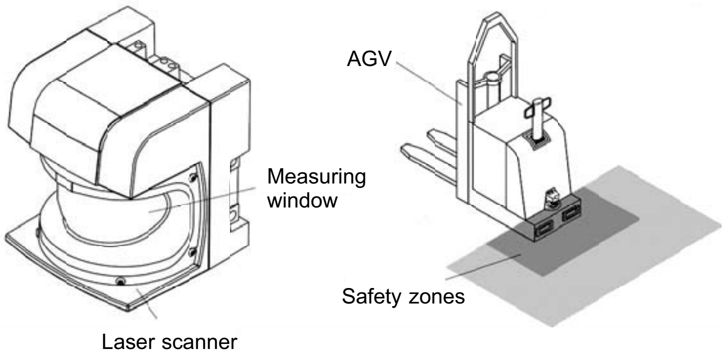


Figure 5.10. Laser scanner for the identification of objects and supervision of the environment

These devices — when stationary — can be used to supervise the machine safety areas by configuring a warning area which overlays the device interfaces. When objects enter the warning field and thus the actual distance falls below the fixed warning distance a signal output is activated at the device which may be connected, for example, with a warning device or an emergency stop at the machine control. Furthermore, this kind of laser scanner is also used on automatic guided transport systems for a continuous environment scan. In combination with a computer the scan field generated by the sensor sets the basis for the vehicle navigation.

During one kind of vehicle navigation the environment fixed coded reflector tags are scanned. A laser sensor mounted on the vehicle or trolley, which is also equipped with rotating mirrors, scans the environment within a range of 360°. The navigation requires only some absolutely scanned measuring points. Based on the topology data stored in the sensor, the horizontal sensor coordinates and, thus, the vehicle coordinates can be determined.

5.3.4 Magnetic and inductive sensors

Magnetic and inductive sensors are used in material flow technology for the contact-free attendance and warehouse control of moving objects and for positioning. Both measuring principles are based on the detection of changes in the magnetic field within the direct scan range of a sensor.

Based on a spindle module, inductive sensors *actively* generate an alternating magnetic field which acts on a receiving coil, which is also integrated in the sensor casing. The voltage induced by this field is amplified in a downstream measuring bridge and directed by an evaluation unit to the sensor outlet.

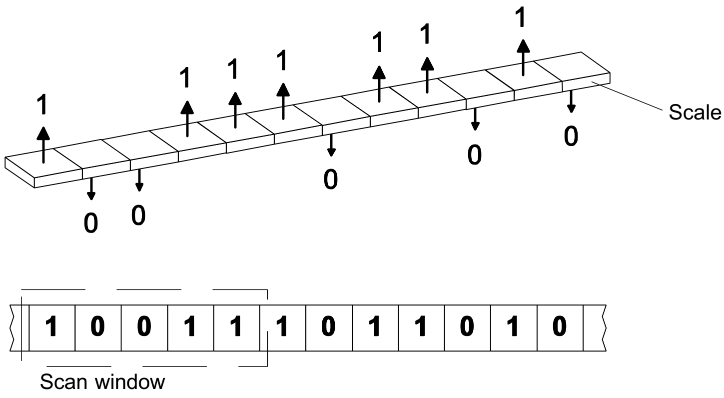


Figure 5.11. Magnetic positioning with an absolute coded scale)

Metallic objects which enter into the scanning range of the sensor change the field (damping) and thus the voltage, which can be detected at the sensor interface as an analog or binary signal.

Inductive *proximity switches* convert the analog voltage at the measuring bridge into a binary output signal with an adjustable switching threshold. The *switching frequency* can thus be varied within strict limits. The typical scanning range of proximity switches is just a few millimeter.

The group of magnetic sensors includes *magneto-resistive sensors* (MR) and *Hall sensors*. Both principles are *passive*, i.e., an external magnetic field is either directly converted into an analog output voltage (Hall principle) or initiates a change of the resistance (MR), which is evaluated in a measuring bridge.

Magnetic sensors are binary transmitters and used in the same way as inductive proximity switches. Objects are detected by means of permanent magnets.

Magnetic positioning systems consist of sensors arranged in a line which scan the magnetic scale at the driving course [34]. The scanning method and the possible resolution are similar to that of optical systems. In this case, the incremental or absolute coded tracks are built by magnetic pole fields on the ferromagnetic base of the scale (cf Fig. 5.11).

5.3.5 Ultrasonic sensors

Ultrasonic sensors are used for the measuring of distances and work according to the runtime measuring principle. The ultrasonic sender in the sensor emits sound waves in a frequency range above the human hearing range (40 up to some 100 kHz) which are reflected by the environment and returned to the receiver.

The measurements are pulsed; the sender generates a sound impulse of a fixed length during each measurement. A downstream controller measures the runtime difference between the generation and the return of the sound wave and stores the measured value in a register. Parameter data can be transmitted, measurements can be initiated and the last measured value can be read via a digital interface. Measured data are put out as time or standardized distance value, according to the parametrization, the read digital value corresponds, for example, to a linear measure in millimeters or inches. Since the values are transmitted by ultrasonic waves the practical average range of a sensor is about 5 m, the measuring accuracy some centimeters. The effects of temperature and humidity on the acoustic velocity of the air have also to be taken into consideration.

Sender and receiver are both mounted in one casing, which may be equipped with a megaphone for the directed sound radiation. It has to be considered that the radiation of the sender is not linearly directed due to overlays and reflections at the megaphone. This leads to additional externally effective secondary areas in addition to the primary club-like measuring areas. This may result in unclear measured values if the sound wave is reflected by several objects at different places in the measuring area.

Ultrasonic sensors are an inexpensive device for environment scans and can be used as anti-collision devices at automatic guided transport vehicles and mobile robots.

5.4 Actuators

Actuator[8] are driving elements which convert the control signals into movements. Actuators, thus, are a link between the information processing on the control level and the automatic material flow processes.

The driving tasks in a warehouse or distribution system reach from the simple movement of a conveyor belt or roller conveyor up to the complex multi-axle movements of a robot. As manifold are the requirements on the actuators (cf. Chapter 4).

The characteristics and specific control features of the actuators are described according to the driving tasks typical in warehouse and distribution systems.

5.4.1 The tasks and structures of actuator systems

There are different kinds of drives [95] depending on the movements of an actuator or its task:

[8] Generally, actuators can be defined as technical systems which convert an electrical input value into an adequate physical output value by means of an emergency current.

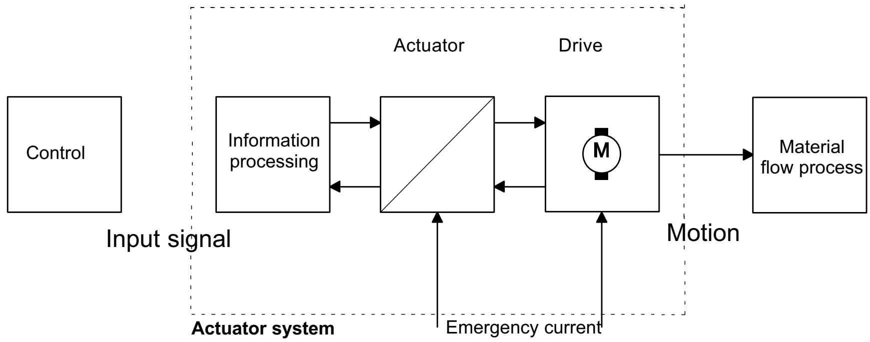


Figure 5.12. Structure of an actuator system

Moving drives are used in most conveying and warehouse work machines, e.g., in cranes, lifts, conveyors and vehicles. They are operated continuously over longer periods of time and described by revolution and torque, force and speed, respectively.

Actuating drives make discontinuous movements to fixed control positions. Such actuators usually approach defined control positions which are determined mechanically or by sensors or by target values specified in the control. Servo drives are used, for example, in discharging modules of sorters,in lifting devices or pallet shifters.

Positioning drives are used as motion drives, for example, in rack feeders or monorail suspension tracks. The control determines a target position along the conveyor line. Unlike in actuating drives the positions can usually be adjusted along the complete conveyor line. The selection of a suitable drive should at first be based on this basic classification. In addition to the actual driving module actuators or *actuator systems* consist of further components which adjust the control signals in an adequate way to the requirements of the drive (cf. Abb. 5.12).

Signal processing
The control values have possibly to be adjusted and converted for a further processing by the actuator. On this input level the information images of the electrical control value (e.g., digital/analogous) are converted and the signal level is adjusted. In non-electrical drives the electrical input signal is also converted into an adequate fluid flow depending on the medium (oil, compressed-air) of the emergency current.

Final control element

The drives are actually controlled by final control elements[9] which also regulate the energy supply. Depending on the emergency current the final element controls the electrical flow or a fluid flow necessary for the operation of a hydraulic or pneumatic drive. In practice, the signal processor and the final control element often built a single unit. Integrated drive controls ensure that the target values (speed, position check) specified by the controls signals are observed by returning the actual driving values. Furthermore, such drive controls supervise the drive unit and in case of an overload can switch off the drive.

Drive unit

The initiation and perpetuation of a movement at a machine require mechanical energy which is generated by the drive unit by means of the supply emergency current and transmitted to the machine. According to the type of the drive unit and the necessary movement translational (lifting cylinder) or rotational drives (motors) have to be used. The mechanical movement generated by a drive unit has to be converted into the required movement for the work machine by means of control shafts or crank drives.

In addition to electrical drives, in some areas of conveyor and warehouse technology also hydraulic and above all pneumatic drives are used (cf. Table 5.3). In this context, special attention should be given to fluid lifting cylinders which are used as control drives in dispatching and sorting plants as well as in palletizers and lifts.

5.4.2 Electrical drives

Since electrical networks are available in all companies and the electrical energy can be stored in accumulators electric motors are often used for stationary and mobile material flow operations.

According to their structure, functioning and operation the electric motors commonly used in material flow technology are classified as follows:

- Alternating field machines (three-phase and alternating current motors)
- Commutator machines (direct current motors)
- Special forms (multiphase motors, linear drives)

The physical specifications of electro-dynamical drives are not described in detail in this book. For further, more detailed information please refer to [52, 57].

[9] In literature this term has different meanings. According to the German standards DIN 19221 or 19226 final control elements here are considered as function elements preceding the actual drive.

Table 5.3. Comparison of electrical, pneumatic and hydraulic drives used in material flow technology (acc. [39])

	Electrical	Pneumatic	Hydraulic
Expenditure for supply of emergency currency	Low	Medium	High
Possible distance	Large	Large	Medium
Movement	Mostly rotational	Mostly translational	Translational and Rotational
Control range	Large	Small	Large
Controllability	Very good	Bad	Very good
Holding torque	Possible	Low	High
Power density	Medium	Low	High
Overload protection	Current limitation	Pressure limitation in compressed-air network, pressure valves at the outlets	Pressure valves
Noise emission	Low	High	Medium
Applications in material flow technology	Universal	Stationary for simple drives(switches, swivel drives)	Stationary and mobile, lifting and driving gears (stacker cranes)

Three-phase asynchronous motors

Three-phase asynchronous motors (TAM) are alternating field machines directly supplied from a three-phase power net. The structure and functioning of such machines as well as the typical methods to adjust the revolution and torque are described below.

At the periphery of the *stator* of a TAM pairwise connected coil halves are mounted on opposite sides between which a magnetic field is generated. The number of coil pairs connected to a phase line determines the number of pole pairs, a bipolar machine hat the number of pole pairs p = 1. Multipolar machines have more pole pairs connected to their stator. Usually, two- or four-polar machines are used with three or six coil pairs (in a three-phase network).

The revolving field is generated by the time mismatch of the phase current by 120° at the connected coils. It induces a flow in the rotor which generates

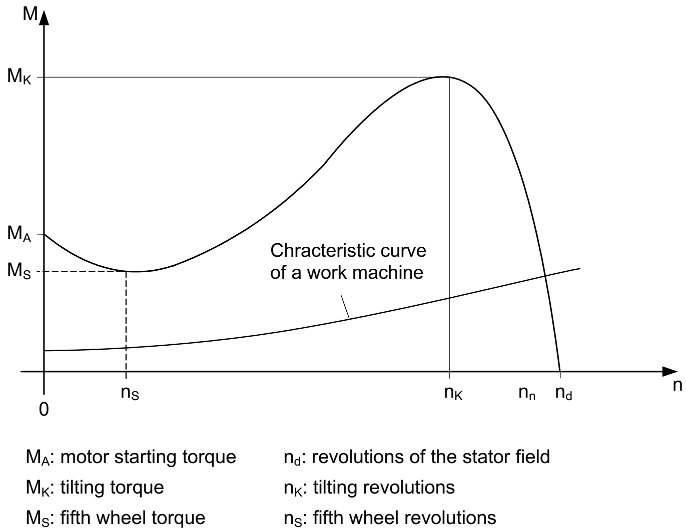


Figure 5.13. Characteristics TAM

a magnetic field with the same rotational direction as that of the outer stator field.

The torque generation in the rotor depends on how fast the magnetic flow between the stator and the rotor changes and is only maintained when there is a slack s between the stator field and the induced rotor field. Only when the machine is unloaded the revolution corresponds to the frequency of the alternating field (synchronous revolution n_D, cf. fig. 5.13). When the machine is loaded the alternating field precedes the rotation of the rotor, thus the term asynchronous machine.

The rotor of a TAM is either a *slip ring rotor* with external connections for the rotor coils or a *squirrel-cage motor*. Squirrel-cage motors or *short-circuit rotors* have no rotor coils, but profile bars made of copper, bronze or aluminium which are connected with rings at both sides of the rotor. Squirrel-cage motors are more robust and require less maintenance than slip ring rotors, which because of their collectors are prone to failure due to burnout or sparks. During start-up a TAM takes up the most energy, about five to seven times the nominal current. To avoid surges in the network, efficient asynchronous motors with a capacity of 4 kW or more have to be equipped with adequate jump starts. If a TAM should be used for RPM-regulated drives additional control measures are necessary which will be described in the following.

The **delta-wye switch** is a simple method to adjust the revolution but can only be used in a fixed revolution range.

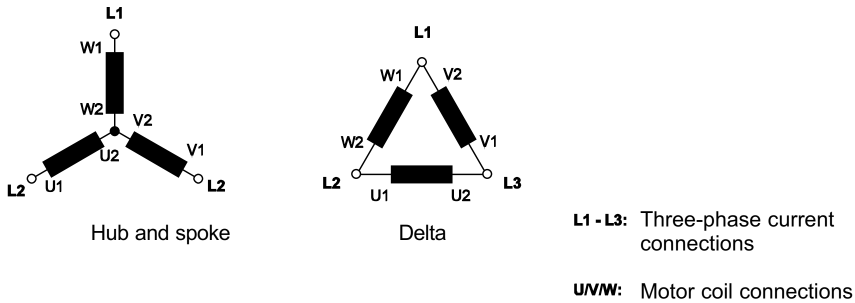


Figure 5.14. Delta-wye switch

When the device is started in the delta-wye modus the coil ends of the stator coil are bypassed at the connections U2, V2 and W2 and the connections U1, V1 and W1 are directly connected to a phase connection of the alternating current network (L1-L2-L3) (cf. fig. 5.14). Thus, two adjacent stator coils are connected in series and because of the higher resistance between the phase connections there is a lower start-up current. After the start-up the machine is switched to delta, the power take-up and the corresponding torque are increased by the factor three. The delta-wye switch is used as a simple jump-starter for motors with a determined nominal revolution (lift drives).

Without an additional circuit a TAM is operated with the revolution determined by the number of pole pairs and the network frequency (50Hz) and has the characteristics shown in fig. 5.13.

In case of a **pole pair switch** motors with a so-called *Dahland winding* are used. Their coils are switched at the periphery of the stator by a *Dahland switch*. In case of four-pole machines (p = 2) either the adjacent or opposite coils can be connected. The revolution of the alternating field then is 3000 min^{-1} or 1500 min^{-1}. Pole switches are only common for squirrel-cage motors. They are also used as drives for lifting devices.

The operating characteristics of a TAM slip ring rotor can be influenced by a **slip control** by installing an additional rotor circuit or by adjusting the voltage.

The power induced in the rotor is weakened by resistors or power conversion cascades in the rotor circuit and either dissipated as power loss to the environment (resistors) or refed into the network. When the tension of supply in the stator circuit is reduced, the start-up torque is reduced over-proportionately ($M \sim U^2$) although this soft-start requires simple start-up conditions (machines with low start-up torque).

The synchronous revolution of the circular stator field and thus the driving revolution can be adjusted infinitely and with low loss by means of a **frequency switch** with a *frequency converter*. Frequency converters are installed between the network and the motor where they rectify the alternating current of the three-phase network and electronically generate a tension of supply with a variable frequency and amplitude at the converter

The number of used frequency converters has increased in recent years owing to the development of ever more efficient and inexpensive devices. The combination of frequency converter and squirrel-cage motors, thus, is a very cheap and multifaceted drive which is more and more used in fields which so far have only used direct current motors. [57]. When they are combined with a frequency converter and high-resolution angle transmitters which are mounted at the drive shaft more and more TAM are used instead of the common direct current motors.

Direct current motors

Direct current motors offer a series of different features:

– Infinitely adjustable revolution over a large range
– Highly stable revolution (little changes of revolution also when loaded and unregulated)
– High synchronization (consistent revolution)
– Highly dynamical (acceleration capability)

Disadvantages are a high wear and tear at the moving current-carrying contacts because of mechanical friction and burnout. Stationary operation furthermore requires a synchronized well-smoothed voltage which has to be generated by additional aggregates.

Direct current motors can be operated as shown in fig. 5.15 by unequally connecting the stator and rotor windings. The functioning of a direct current motor is mainly influenced by the type of excitation and classified as follows.

In **switch motors** the armature and field windings are arranged in line and operated with the same voltage.

The torque characteristics of the switch motor are calculated from the armature voltage U_A, the resistance of the armature circuit R_A as well as the machine invariables c_M according to eq. 5.1 into

$$M = \frac{c_m \cdot \phi \cdot U_A}{R_A} - \frac{(c_m \cdot \phi)^2 \cdot n}{R_A \cdot 2 \cdot \pi} \tag{5.1}$$

In case of an invariable flow ϕ and a constant armature voltage U_A the characteristic is linear.

The start-up behavior is influenced by an adjustable started switched in series to the armature which can be bypassed during operation. When the values of the starter resistor are lower the characteristic curves flatten so that the start-up torque and the start-up power can be reduced. The field winding also passes the starter so that the self-induction voltage is bypassed via the starter resistor and the armature circuit when the machine is switched off. A disadvantage is the low efficiency due to the power loss in the starter control.

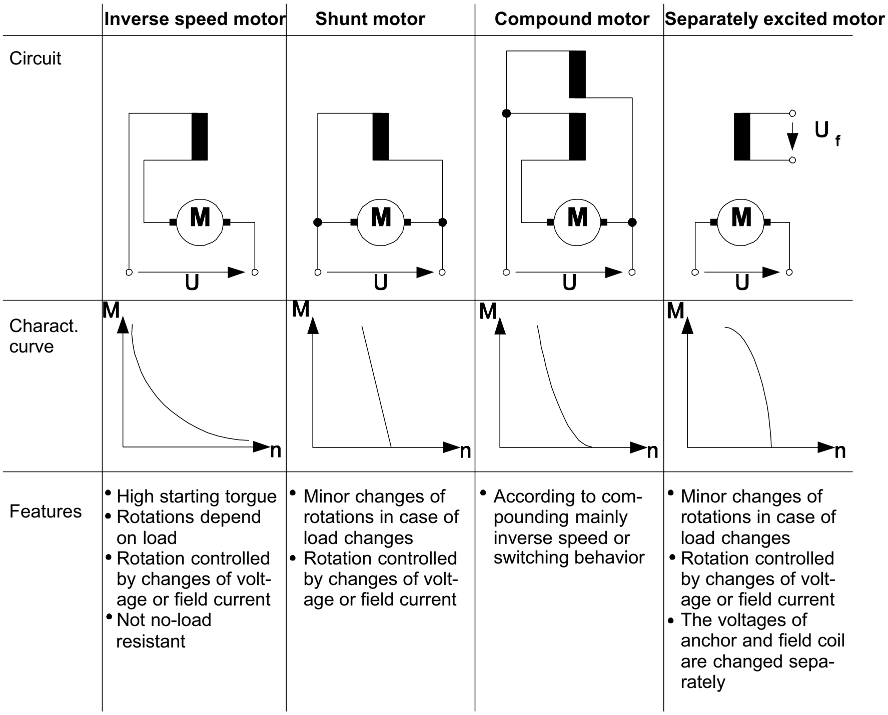


Figure 5.15. Excitations of direct current motors

In a **series motor** the field and armature winding are switched in series and are thus operated with a direct current.

Unlike in a switch motor revolution and torque are not linear because of the electric circuit (cf. fig. 5.15). This motor is marked by an easily adjustable revolution and a high start-up torque. It is mainly used in lifts and drives at unsteady conveyors.

In view of their characteristic curve series motors must not be operated unloaded or equipped with just a revolution control. When the load moment sinks the revolution may rise to a critical value and the resulting centrifugal force may lead to a "runaway".

The revolution can be controlled by a *field controller* arranged parallel to the field winding or by tapping the field winding, i.e., by reducing the number of windings.

The **compound motor** has a series winding in addition to the switch field winding. If both windings are wound in the same direction, the revolution behavior is between that of a switch motor and that of a series motor. Revolution and torque are influenced similar to a series of switch motor. The term *compounding* (cf. Fig. 5.15) describes the design and use of the field winding.

Separately excited motors require separate power supplies for the armature and field circuit. The mechanical structure corresponds to that of a switch motor. Since armature and field circuit are separated the exciter field remains constant when the voltage is reduced. Thus, the revolution is more stable than in a switch motor (more stable drive). The revolution is proportionate to the adjacent armature voltage.

For smaller motors permanent magnets are used instead of an electric separate excitation. These motors are used, for example, as final control elements for valves or as servo motors.

5.4.3 Fluid drives

Hydraulic and pneumatic drives are used as driving and control elements in conveyors and warehouse facilities. Although electric drives can also be used for these tasks and are more common fluid actuators are more economical and functional for certain applications because of their compact construction. The function elements of fluid drives are:

– Pressure generators, pumps (hydraulic) or compressors and compressed-air containers (pneumatic)
– Drives (lift cylinders, axial, radial piston engines)
– Final control elements (valves to control the fluid flow)
– Wires and pipes to transport the medium (oil, air)

The compressed-air for pneumatic plants can be provided by a central compressor which supplies all decentralized outlets throughout the company via pipelines. On the other hand, the compression for hydraulic systems has to be generated in the direct vicinity to the drive for safety reasons and to return the medium into the closed cycle.

In stationary facilities electrical compressors are used while hydraulic vehicle are operated with electric as well as combustion engines. Hydraulic drives are characterized by a high power density (transmission of large forces within a small area) and an excellent drive stability.

Typical applications for translational *hydraulic drives* are scissor-lift tables and lifts as well as lifting devices at fork-lift trucks. Hydraulic motors are also used in very small rotation engines in stacker drives (hub motors). Final control elements in fluid systems are mechanically and mainly electrically operated distributing and control valves which control the fluid flow and limit or regulate the flow and pressure. The main characteristics of hydraulic systems are described in Table 5.4.

Pneumatic drives are mainly used as final control elements in swivel devices, pushers or stoppers in conveyor belts and in sorting and distributing facilities. The required capacity is lower and no high requirements are put

Table 5.4. Characteristics of hydraulic drive systems

Advantages	Disadvantages
– Generation and transmission of large forces (pressure range typ. 200 to 500 bar) – Infinite adjustment of working speed – Single reversion of moving direction – Simple control by automated systems (electrically operated control and distributing valves)	– Additional aggregates for filtering of oil – Short transport distances – Prone to leakages due to high pressure – Return required (closed cycle)

Table 5.5. Characteristics of pneumatic drives

Advantages	Disadvantages
– Long transmission distances possible (central generation of compressed-air in the company) – Central overload protection (compressed-air limitation) – Possibility to store the air – No return required (air is blown out) – Simple control by automatic systems (electric distributing valves)	– Compressability of the air, no uniform movement – Only small force transmission possible (pressure range 6 to 10 bar) – Noise emission – Condensation

with regard to the precision of the control movements. Table 5.5 shows the basic characteristics of pneumatic systems.

For further more detailed information about fluid drives please refer to [13, 104].

5.5 Interfaces in automation systems

Computer systems sensors, actuators and periphery components have to be linked by suitable interfaces to ensure a smooth data exchange. According to the presentation of the information of a transmitted signal, the point-to-point interfaces can basically be divided into binary and digital interfaces. The digital data transmission is, furthermore, divided into a serial and a parallel transmission.

An important criterion for the selection of an interface is the quantity and speed of the data to be transmitted.

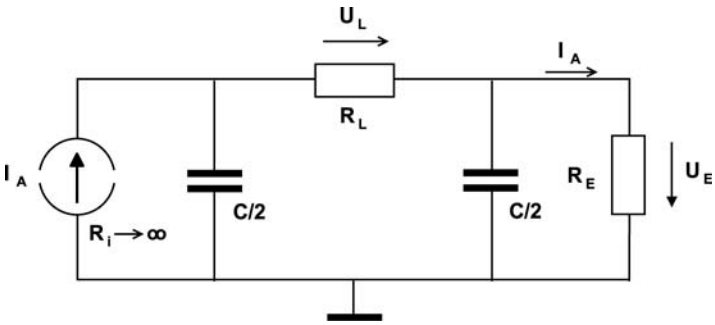


Figure 5.16. Equivalent circuit diagram for a current interface

While large data quantities of up to some Mbytes are transmitted between ERP systems and master computers (cf. Chapter 7) only small data quantities of only a few bytes are transmitted on the field level between sensor, actuators and a process-oriented control. These data, however, have to be transmitted in realtime to ensure that messages about rapid status changes can be sent to the control without delay.

Based on the variety of communication requirements, different types of interfaces have been established which are each used for special applications. In the following interfaces common in subordinate automation technologies in warehouse management systems their characteristics and fields of applications are described.

5.5.1 Analogous and binary data transmission

The classical and technically most advanced method to connect field-oriented components is the use of a two-position or parallel connection where all sensor or actuators are connected star-shaped with the control by a pair of unscreened cores each.

Via this interface analogous as well as binary signals can continuously be transmitted between two terminals. In a *current interface*[10] (cf. Fig. 5.16) an analogous signal is standardized to a signal level of 0 or 4 to 20 mA and to 0V and 10V for *voltage interfaces.*

When current interfaces are used the analogous signal is not tampered by ohm losses caused by the output resistance. This is the case with voltage interfaces so that without an additional amplification the transmission distances are limited to some meters.

The output signal of the sensors is adapted to the signal range of the respective interface and is converted by a PLC or process computer in the input module into a digital variable for further processing.

[10] The most common terms are TTY or current loop.

At this interface mainly binary signal transmitters and actuators are connected with a superordinate control (PLC). No special requirements are put on the wires because the binary transmission with a usual voltage of 0V and 24V is relatively failure-free also over larger distances thanks to the switching tolerance of a PLC (cf. 5.2.2).

A disadvantage, above all in large plants with many inputs and outputs, is the extensive wiring since each device needs its own pair of cables. For this reason the two-position connection is mainly used for the local connection of a limited n umber of sensors and actuators to decentralized controllers or connection modules of field busses.

5.5.2 Digital data transmission

Serial and parallel interfaces are the basic tools for the digital data transmission between computer systems and periphery or automation components. Both methods are presented in the following together with the prevalent industrial standards and typical characteristics (transmission rate, length of wires, interface design).

Parallel interface

Many PCs offer the possibility to control periphery devices via a parallel interface and to transmit data byte by byte. Other than serial interface parallel interfaces (Table 5.7) allow for a high data transmission rate. Due to the relatively long wires and the high expenditure these interfaces are only used to connect local devices, above all printers [58].

The connection between a PC and a printer consists of a 25-pole Centronics cable[11]. In addition to the eight data lines (*D0* - *D7*) and the signal mass (GND) additional poles are needed for the so-called hardware handshake which controls the communication between the two devices. The respective control lines are set at both sides of the sender and receiver to show that data are available at the interface (BSY) and can be taken over by the receiver. Furthermore, the receivers acknowledge if the printer has successfully received data ($\overline{ACK}$ - Acknowledge[12]). Further control lines show the status of the printer (*PE* - Paper End, *ONL* - Online) or reset both communication partners to a defined initial status ($\overline{RESET}$). The voltage is between 0V and 5 V.

Serial interfaces

During a serial data transmission the digital information is transmitted sequentially bit by bit. Instead of eight or more data line for a parallel,

---

[11] Another form with higher transmission rates in the SCSI standard which is also used to connect external devices but requires additional hardware (controller).

[12] The signal description shows that the signal is activated in case of a falling edge, e.g., at a signal voltage of 0V.

Table 5.6. Characteristics of the Centronics interface

Transmission rate	Up to 100 kByte/s
Wire length	Typ. 5 m (up to 15 m for low transmission rates), special solutions of up to 400 m with 20 kByte/s
Connection type	25 pole Sub-D, Centronics (36 pole)
Connection	Point-to-point

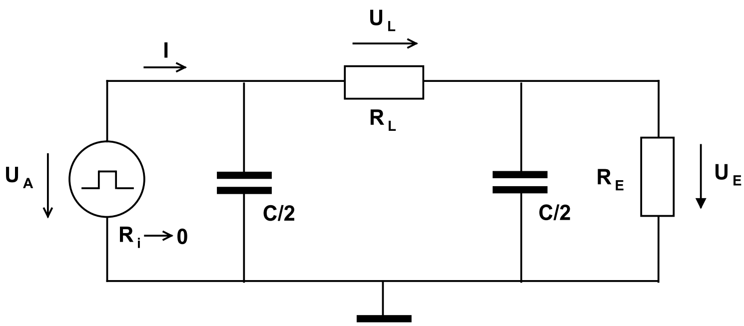


Figure 5.17. Equivalent circuit diagram of a voltage interface

character-by-character transmission, only one line or pole pair is necessary, depending on the design. Thus the wiring expenditure is generally smaller.

The standard RS232 specifies the physical interface for a serial data exchange between two terminals as a point-to-point connection. This kind of interface requires one data cable for each transmission direction. For a full duplex transmission (cf. Chapter 7) where both users can send and receive simultaneously at least three poles (transmission, receiving and ground line) have to be connected.

The signals are transmitted via a bipolar voltage interface with a typical signal level of +3V to +12V or "High" to "Low" compared to the ground potential. Due to the ground connection, the line is more sensitive to external electro-magnetic disturbances. Furthermore, the ohm and capacity losses reduce the maximum transmission rate and the length of the wires. In case of a digital *serial* data transmission, the same electro-technical conditions occur as during a binary or analogous signal transmission. The line capacity which in this version together with the line resistance R_L builds an RC-link inevitably influences the signal transmission and thus limits the maximum transmission rate (cf. Fig. 5.17).

RS232-interfaces are mainly used to connect PC periphery components with only little requirements on the interference resistance and data integrity,

Table 5.7. Characteristics of serial interfaces

	RS232	**RS422/RS485**
Transmission rate	Max. 115kBit/s, typ. 9600Bit/s	100kBit/s
Wire length	Typ. 15-30m (9600Bit/s)	Up to 1.200m at 100kBit/s
Connection	Two users (point-to-point)	RS422: up to 10 users RS485: up to 32 users
Connection type	Standard 25 poles Sub-D, also nine poles	Sub-D, nine poles

e.g., printers, handheld scanners, as well as as programming interfaces for automation devices.

Universal serial bus USB(USB) is another industrial standard for serial interfaces for the connection of periphery devices, which unlike RS232 are bus-compatible.

Compared to RS232, the USB has a higher transmission rate, similar to that of the Ethernet[13] and offers the great advantage that periphery devices can be plugged in and out while the PC is running. Another advantage is that the supply voltage is also fed into the bus so that many USB devices do not need a separate power supply. With regard to the short wire lengths, however, USB can only be used to connect local computer peripheries.

The interfaces R422 and RS485 transmit the data by means of floating voltage signals. A two-wire line for each channel is included in the cable (Fig. 5.18). Four data lines are required for a full duplex transmission. Thanks to their physical transmission principles, both interfaces are failure-free and can thus be operated with smaller signal levels and higher transmission rates[14]In the two-wire version RS485 is the most widespread standard for *field bus systems.*

5.5.3 Field bus systems

On the control and field level field bus systems are used to connect field devices (sensors, actuators) with the field-oriented control.

Field busses are used for several reasons:

– To reduce the expenditure for the planning, projection and realization compared to a star-shaped single connection of field components
– To use a standardized data communication for all users

[13] The USB standard 1.1 provides for 1.5 or 12MBit/s, USB 2.0 for a maximum of 480MBit/s.

[14] The most important factor is the time for the growth of the signal edge which because of the lower voltage level is shorter.

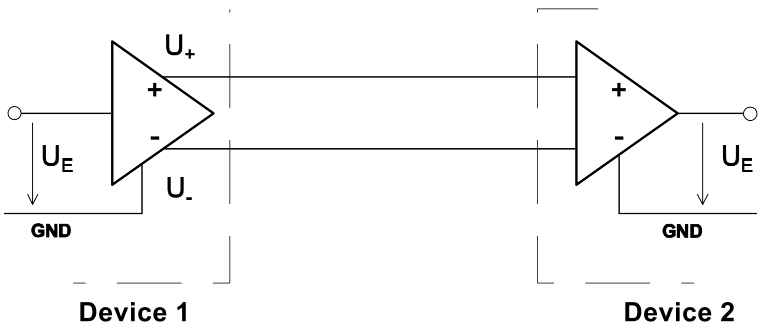


Figure 5.18. Symmetrical floating voltage interface

– To reduce the expenditure for the software implementation by uniform interfaces and telegrams
– To render the digital and above all the analogous I/O modules in the control unnecessary
– To ensure a higher flexibility towards changes and reconstructions

In field busses data can be transmitted via an electrical line (typ. two-wire line), a light conductor or a wireless transmission channel. Electric two-wire lines according to IEE 802.3 (Ethernet), RS485 or IEC 1158-2 are the most popular interfaces in the automation of production and material flow.

Optical wave guides are mainly used in explosive areas in process engineering because of their high failure resistance and intrinsic safety. An electric intrinsically safe version of IEC 1158-2 has been specified with a reduced signal voltage and currents and a maximum transmission rate as to avoid an ignition of the environmental mixture of gases.

Busses are designed as direct serial connections between the users and according to the topology they may be a bus (line), tree or ring. While the users communicate, data are exchanged by telegrams where the actual user data and ID are stored. The connection structure ensures that all users are connected to the bus at the same time and principally can send data simultaneously. This would inevitably lead to a communication breakdown. To keep the data exchange failure-free regulations have to be set up which control the access on the bus:

CSMA/CD - Carrier sense multiple access / collision detect This is a decentralized access method which principally allows each user to send a telegram at any time. The users check before each transmission if other stations use the bus (CS-MA). If no communication takes place a telegram can be sent. Because of the signal runtimes, which in large networks can be very long, two different users may send two telegrams at the same time.

Should the data collide the transmission is cancelled and the telegram is resent at random (CD).

CSMA/CD was designed as an access method for busses with many users (e.g., Ethernet). The busses are configured for a clear user address. Due to the undeterminate access such a bus is not realtime compatible.

Master-slave describes a method where a user (master) centrally controls the communication with the bus and grants transmission rights to to other users (slaves). All slaves are *cyclically* addressed by the master, the access frequency of single slaves can be parameterized in the bus master.

Thus, each user has access onto the bus in a fixed order and for a defined period of time. This fulfills the requirement for a realtime capability. It is a disadvantage, however, that no communication is possible when the bus master breaks down. The data transmission between two slaves is more complex since it is done in two telegrams via the master. The cycle time increases in line with the number of users so that the latter is limited. Furthermore, each time a user is added or removed the master has to be reconfigured.

Token-passing is a communication method where a special telegram (token) is exchanged in turn between all users. When a user receives the token he is granted the transmission rights and may send a data telegram to the bus. After a defined period of time the token is handed over to the next user.

Since the token retention period is defined by the controller the response time per cycle is predetermined. The users can communicate directly with each other and there is no communication overhead as in the master-slave method. Here as well, the cycle time increases in line with the number of users and the bus has also to be reconfigured when users are added or removed.

In case of the **sum frame method** which in Interbus-S controls the access onto the bus the process and parameter registers of all bus users are represented by a so-called **sum frame**, which cyclically circulates in the bus. The bus is a physical ring line with two pole pairs. Each user functions like a shift register which receives a byte from the neighbouring user via an interface connection, moves the internal register contents by one location and sends the removed byte to the next user from the other interface connection. Additional bits in the data flow hold a code which bytes on the bus are read or written by the users. A cycle is thus defined by the number of bytes and control bits in the **sum frame** and the determined data rate.

This transmission method is very efficient thanks to the telegram overheads. In addition to this, the consistency of the data in the sum frame is automatically ensured by the procedure. However, several cycles are required to transmit larger data blocks. As with the master-slave method no communication is possible when the master breaks down. In this bus configuration

Table 5.8. Comparison of the characteristics of different field bus systems

Bus system (Provider)	**Number of users without (with) repeater**	**Transmission rate, maximum wire length**	**Bus access method**
Ethernet	100 (open)	100 MBit/s, 500 m	CSMA/CD
PROFIBUS-FMS (Siemens)	32 (127)	93.75 kBit/s, 500 m ⋮ 1.5 MBit/s, 200 m	
PROFIBUS-DP	32 (126)	9.6 kBit/s, 1200 m ⋮ 12 MBit/s, 100 m	Token-passing combined with master-slave
PROFIBUS-PA	32 (126)	31.25 kBit/s, 1900 m	
CAN (Bosch)	32 (open)	10 kBit/s, 5000 m ⋮ 1 MBit/s, 25 m	CSMA/CD
LON (Echelon)	32 (open)	39 kBit/s, 1200 m ⋮ 1.25 MBit/s, 20 m	CSMA/CD
Interbus-S (Phoenix Contact)	256 (256)	500 kBit/s, 12800 m	Sum frame
ASI (Siemens, Pepperl+Fuchs)	31 (31)	167 kBit/s, 100 m	Master-slave

this is also the case when a slave has broken down. Here as well, the bus has to be reconfigured when new users are added.

For a *field bus implementation* (cf. Table 5.8) the implementations of the bus communication methods of the suppliers are defined.

Because of the large variety of different products, there is no uniform standard for field bus systems. For this reason, the bus systems of different providers are compatible only under certain circumstances. But the busses of different providers can be connected by means of additional *bus converters*.

The *Actuator-sensor interface* (ASI[15]) plays a special role because it is no field bus in the narrower sense, but was developed especially for the connection of binary sensors and actuators. In addition to binary data, the bus also transmits the emergency current for the power supply of the connected field devices; four binary field devices each can be connected to the bus by a maximum of eight *coupling modules*. The number of users for each bus segment is limited to 31, the slave address 0 is not assigned.

Independent of the telegram structure the number of users in the bus is physically limited by the electrical line characteristics and by the power consumption of the slaves. To increase the number of users repeaters are used as amplifiers but not in field busses with a fixed telegram structure (e.g., interbus and ASI).

Since field busses are used for a large range of applications the suitable choice largely depends on the concrete communication requirements. A general statement about the suitability of a certain bus system cannot be given. Busses working according to the token-passing procedure should preferred for realtime critical applications because they have a deterministic response time behaviour because of the predictable token runtime. In these systems, the breakdown of the master causes almost no problems because each other user may also be the master.

[15] The term actuator-sensor interface is also used.

6. Automatic Identification

The immediate automatic identification of objects (*auto-ID*) for the management and control of processes in production and logistics is a basic element of each WMS. The immediate faultless identification of objects is an elementary condition for an efficient process for which automatic identification (auto-ID) is the basis.

The tasks to be fulfilled in a warehouse put different requirements on the information technology and the choice of an adequate code. These are among:

- An ensured *reading reliability* under the given conditions
- A sufficient *reading speed* for existing and planned conveyor techniques
- The possibility to generate a sufficient number of *identification tags*.
- A *reading distance* adjustable to the given conditions.
- An ensured *compatibility* to other supply chain members.
- The *cost efficiency* of identification systems and operating means.

The choice of a suitable identification system requires an understanding of the basic auto-ID principles and techniques which are studied in the following. In more than 70% of auto-ID application the barcode is the most commonly used method and therefore it is given special attention. But the promising transponder technology will be discussed as well.

6.1 Codes and characters

In the German standard DIN 44300 part 2 the Deutsche Institut für Normung deals with the presentation of information and the related terminology and defined codes:

> "A regulation for the clear assignment (encoding) of characters out of a character set (original set) to those of another character set (image set). Note: The assignment is not necessarily bijective." [19]

Thus, a code is a character set for the representation of original sets by image sets during the encoding including the rule for its presentation [7]. Encoding means the conversion of one character set into another. Thus, a character is:

> "An element (type) out of a finite quantity of objects defined for the presentation of information (character set) as well as each image (copy) of such an element." [19]

It is explicitly differentiated between the element and its image or the type and the copy. Therefore, a character as an object out of the character set may appear several times in the image.

6.1.1 Encoding

Encoding can be described mathematically as an image C which assigns elements of the quantity A (characters of the original set, source alphabet) to elements of the quantity B.

$$c : A \rightarrow B$$

The characters of the original set can be combined into character chains and thus words of the target alphabet. If the image C is bijective this is also called the decodability of the code.

It has to be pointed out, however, that the decodability of character chains of a target alphabet B built of the image C is not always desired. The passwords of an operating system (UNIX), for example, are encoded in such a way as to avoid a clear decoding. Especially in case of passwords, it is important to compare the words with the image set.

6.1.2 Encoding examples

By means of the character set of the alphabet A with $A = \{a, b, c, \ldots, z\}$ the single characters can be encoded into another system, e.g., into Braille or embossed printing. Braille translates each character into 3×2-matrixes $M^{3,2}$ with the elements m_{ij}, with $m_{ij} \in \{0, 1\}$ and $i \in \{1, 2, 3\}$ as well as $j \in \{1, 2\}$.

A blind person cannot see the elements 0 and 1; therefore, these signs are embossed on a sheet of paper. Since the number of elements in Braille is larger than the number of characters in the alphabet[1] and the characters are clearly assigned to a dot combination Braille texts can be decoded.

Another well-known encoding method is the Morse alphabet where the characters are translated into our alphabet by sounds[2]. The target set actually consists of only four signs: a short and a long sound as well as a short

[1] Actually there are more Braille characters than characters in the alphabet if upper and lower cases are not considered. But some dot combinations are used for special characters like full stops, asterisks, exclamation marks, etc. but also to build syllables and thus to improve the readability.

[2] In everyday life quite a variety of encodings can be found, e.g., the encoding of music data [26] which can be read by man and machine thanks to different notations.

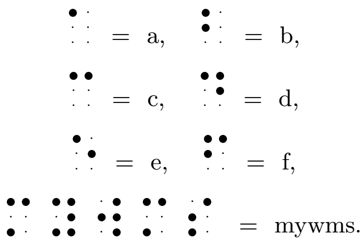


Figure 6.1. Example of a character encoding with Braille

and a long pause between the single sounds and a still longer pause between the single characters. Without the longer pause the single Morse characters could not be differentiated.

In contrast to Braille, which builds alphabetic characters with different combinations of six dots, the Morse code is error prone because it allows for the generation of sound sequences with different numbers of sounds. Messages are decoded for different reasons. In this case, however, it is advisable to give special consideration to the image and target set, the purpose and field of application, the density of information and the error rate prior to developing a new code but also prior to choosing an already existing encoding system.

6.2 1D–Codes

If the symbols of the Morse code for short and long sounds, i.e., the dots and lines, were heightened and the pauses were presented as spaces this would resemble a barcode. But this barcode would be difficult to read by machines and — as described earlier — would be error-prone. If, for example, an *e* consists of a short and a *t* of a long sound the *a* is built of a long and a short sound. Because of the high misinterpretation rate in case of incorrect pauses between the single sounds and the combined sounds such an encoding method is inadequate.

Telegraphy is an old technique. When on 4 September 1837 the first Morse telegram was sent, nobody thought about barcodes. It was not until 1949 when the barcode was patented[3] setting the basis for the automatic identification.

Codes do not depend on their physical presentation. Typically, Morse signs are read acoustically, Braille signs tactilely and barcodes optically[4].

[3] On 20 October 1949 Norman J. Woodland and Bernard Silver presented their invention which on 7 October 1952 got the patent number 2612994 and set the basis for the barcode [81].

[4] In case of bad environmental conditions, other physical presentations such as slotted plates are used which are not scanned and read optically but inductively.

Here, the light reflected by the barcode label is recorded and converted into electric signals.

With the advancement of opto-electronics in the 1950s and 1960s and the call for a quicker identification of objects, triggered by the growing queues at the cash points in self-service shops and supermarkets [33], the code 2/5 (also 2 out of 5) was introduced for the first time in 1968. This type of barcode is still used today[5].

6.2.1 Code 2/5

First of all the different barcode symbols have to be classified into *double-width* and *multi-width codes*. Some symbols use only two different widths for their lines and one constant width for the spaces Fig. 6.3). More advanced multi-width codes use more than two different widths. The code 128, for example, works with four different widths for lines and spaces (cf. Fig. 6.9). Other differences are:

- the sign intervals (e.g., numerical or alpha-numerical)
- the interleaving (the use of spaces for the transmission of information cf. section 6.2.3)

Non-overlapping double-width codes or barcodes of the first generation, like the 2/5 codes are a simple graphical binary image. Digits are coded binary with lines of two different widths (a narrow line for a zero and a wide one for a one). Each digit is built of five lines: two wide and three narrow ones. When systems with five elements are built by permuting two elements of a set, whereas one element has to appear twice in the new image, there are exactly $\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$ (combinations without repetition) possible systems. Thus, the ten digits 0...9 can be encoded.

The barcode has a defined start and a defined end. It is not enough to print a number of bars one after the other to represent the single digits, but a stop code has to be set at the start and the end.

Since with 2 out of 5, the symbol set is exhausted by the presentation of the digits $0 \ldots 9$. The S start and end symbol 2 out of 3, consisting of two wide and one narrow bar, are used (cf. Fig. 6.2).

Start and end symbols are used as start as well as stop characters for the scan process. Barcode scanners work independently of the orientation of the barcode. To ensure a secure scan *all* bars have to be scanned independently of the reading direction. Each set of barcode symbols requires a unique pair of start and stop symbols. Furthermore, the trigger helps to identify the barcode symbologies.

During its lifetime an object may have been "labelled" with several barcodes. In case of an insufficient coordination, for example, the companies

[5] The code 2/5 was developed by Identicon Corporation.

Table 6.1. Code 2/5

Character	Binary pattern	Barcode	Character	Binary pattern	Barcode
Start	110		Stop	101	
0	00110		1	10001	
2	01001		3	11000	
4	00101		5	10100	
6	01100		7	00011	
8	10010		9	01010	

= Start = Stop

Figure 6.2. Start and stop symbols of the code 2/5

participating in the supply chain possibly use internal barcodes. If the code belongs to a different symbology, this is recognized by the scanner and transmitted selectively. This helps to reduce the number of faulty identifications and improves the correct assignment. Generally, the scanners can be configured for the selective transmission of certain barcode symbologies.

An important feature of all barcodes is the code length, and thus the space it requires when printed. The different barcode symbologies meet the requirement for a maximum of information on a minimum of space in different ways.

The most narrow element in a barcode (the narrow line) is also called *module* and its width is the width of this most narrow element. In the code 2/5 the width of the wide bar S_f usually is trice as wide as the module, i.e., as the narrow bar which in the following is called S_d. The blank S_z between two bars has the double module width. The length L_z of a digit code can now easily be determined:

$$L_z = 2S_f + 3S_d + 5S_z = 19S_d$$

Thus, the length of a digit code is 19 times the module width. Five spaces are needed to separate the five bars because the next digit code also starts with a black bar. Analogously the start character can be determined by $L_s = 13S_d$ while the stop character is determined by $L_e = 11S_d$. L_e is clearly shorter since unlike L_s L_e is not followed by another character.

Each barcode has to be preceded and followed by a space to avoid faulty scans. This space is called *quiet zone* R_z; the quiet zone has to be ten times as wide as the module, but at least 2.5 mm (cf. [70]): $R_z = 10S_d$. The total encoding for n digits has a length L_{total} of:

$$L_{ges} = 2R_z + L_s + nL_z + L_e$$

Based on the concrete module width the absolute length of the barcode can be determined.

Table 6.1 shows the symbols for code 2/5. If, for example, the digits 4465 should be encoded with the code 2/5 together with the start and stop code this would result in the concatenation[6] of the combinations of bars (in Table 6.1) as is shown in Fig. 6.3.

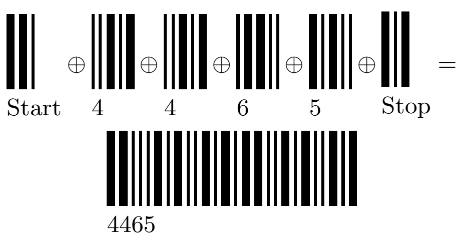


Figure 6.3. The digits 4465 in code 2/5

The characters of code 2/5 shown in Table 6.1 and the system described above have a high error detection rate. According to the regulation, a single digit has to consist of five bars, two wide and three narrow ones.

Fig. 6.4 shows a faulty barcode for the digit 4. The results differ depending on how the scanner reads this code: Scan A actually identifies the digit 4 with the binary pattern 00101 while scan C identifies the binary pattern 10001 and thus the digit 1. Scan B, on the other hand, detects an error in the binary pattern 10101 because there are exactly two wide bars in each digit code. However, is is quite unlikely that such an error would occur, i.e., that in a case like in Fig. 6.4 the digit 1 would be read because there would have to be

[6] In the following the concatenation, i.e., the chaining of characters and character chains is marked by the operator $\oplus$. The concatenation is not to be confused with the addition $+$.

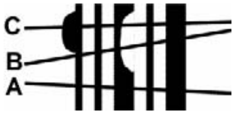


Figure 6.4. Code 2/5 partially with errors in the printing image. A,B,C: Scan lines

two errors in one character which would result in a readable character with two wide and three narrow bars[7].
There are three different scan methods:

– The correct scan of the code (A in Fig. 6.4).
– The faulty scan which is signalled by a corresponding error message by the scanner(B in Fig. 6.4).
– The unrecognized faulty scan "WrongRead" which returns an image which does not correspond to the actual code (C in Fig. 6.4).
– The "NoRead" i.e., the message that something was scanned but could not be identified
– The NoRead that is not acknowledged by the hardware.

This clearly shows the advantages of the code 2/5, compared to the above-mentioned presentation of characters in the Morse alphabet. The following section shows that the error detection in a barcode and the safety of this system can still be improved considerably.

6.2.2 Check digit calculation Code 2/5

The occurrence of a *substitution error*, i.e., the opposed alternation of two bar widths in a digit barcode is highly unlikely but has still to be expected. This error can further be eliminated by a check digit as the last character before the stop character. With this simple but efficient method the unrecognized scan of a substitution error can be reduced by more than 50 percent:

Starting at the right, one can add up the active digits c_i of the digit to be encoded are multiplied alternately by three and one and results: $s = 3 \times c_1 + 1 \times c_2 + 3 \times c_3 + 1 \times c_4 + 3 \times c_5 + \ldots$. The sum s is calculated *modulo* ten[8] $a = s \, mod \, 10$ and the result a is subtracted from ten. The resulting c_0 corresponds to the check digit. However, if $c_0 = 10$ c_0 is set to $c_0 = 0$:

$$c_0 = \left(10 - \left(\left(\sum_{i=1}^{n} c_i \times \big(1 + (i \, mod \, 2) \times 2\big)\right) mod \, 10\right)\right) mod \, 10$$

[7] Environmental impacts affect the optical and geometrical properties of the barcode label. Different processes have been developed to improve the scan quality. A promising solution is the use of fuzzy barcode decoders in the form of software [8].

[8] Whereas the operation modulo corresponds to the remaining value after division by a number. Example: $17 \, mod \, 5$ because $17/2 = 3$.

According to this formula the check digit for 124 is $(10 - ((4 \times 3 + 2 \times 1 + 1 \times 3)\, mod\, 10))\, mod\, 10 = 3$.

An example of a complete and correct barcode 2/5 with the digits 124 and the check digit 3 is shown in Fig. 6.5.

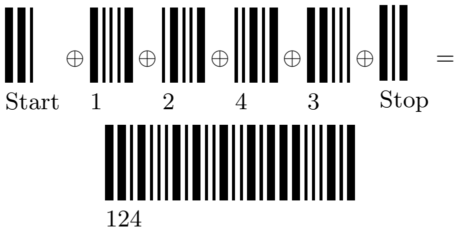


Figure 6.5. Digits 124 in code 2/5 with check digit

The generation of a check digit is obligatory because it is a regular part of the code 2/5. A missing check digit would lead to a recognizable error or an unrecognized wrong read. The number 31509, for example, can be encoded as 315098 after generation of the check digit. If this check digit is missing the number 31509 is decoded as 3150 because 9 is the check digit of the number 3150, while 0 is the check digit of 315, 5 the check digit of 31 and 1 the check digit of 3. There is a possibility of 10% for each combination of digits that the last digit is a valid check digit. To reduce faulty or wrong reads is one aspect which has to be considered when choosing barcode symbologies for the automatic identification. If the scan rate in a warehouse, for example, is 10.000 units/hour an error rate of one-tenth of a percent would lead to an average of ten events which typically would require manual correction. Depending on where in the warehouse this error occurs, it would need additional manpower to correct it and the consolidation of orders would fail.

The choice of the right barcode symbology and the correct technical print and scan is a decisive factor. Furthermore, a warehouse planning always should consider the question of what would happen in case of a wrong read of a barcode.

6.2.3 Code 2/5 interleaved

Based on code 2/5 in 1972 the code 2/5 interleaved was introduced. The code 2/5 interleaved can be seen as a refinement of the conventional code 2/5 because the spaces between the black bars are used for the information transfer so that the information density is much higher. The spaces are used as shown in Table 6.1 whereas the codes are included in the white spaces instead of in the black bars. Accordingly, these white spaces have different widths, three times narrow and two times wide. For the code of the number

z (including the check digit) with $z = c_n \oplus c_{n-1} \oplus \ldots \oplus c_1 \oplus c_0$ it applies that each c_i with an even index i is presented by a black bar while each c_i with an odd index i is presented by a white bar.

Since the code 2/5 interleaved represents a barcode symbology of its own, it gets a new start and stop character (cf. Fig. 6.6).

Figure 6.6. Start and stop symbol of the code 2/5 interleaved

The following applies to the start symbol: It consists of four narrow bars (two black and two white)[9] in the sequence black, white, black and white, $L_s = 4S_d$. The stop symbol consists of one wide black bar followed by a white narrow one and is terminated by a black wide bar; the length is $L_e = S_f + S_d + S_d = 3S_d + 2S_d = 5S_d$. This makes the code 2/5 interleaved much more compact.

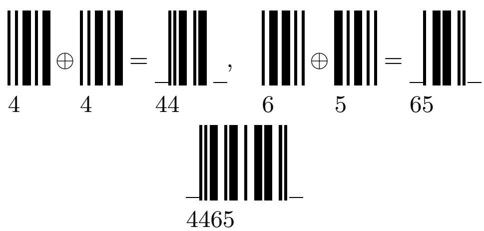


Figure 6.7. The number 4465 in code 2/5 interleaved without start and stop character and without check digit

Fig. 6.7 shows the structure of the code 2/5 interleaved: Instead of a fixed width for the spaces these have the same width as the bar of the following symbol.

As can easily be seen in the code 2/5 interleaved only odd numbers of digits can be encoded so that the number of bars will be even when the check digit has been added. When the number of digits is even the set has to start with a zero. The encoding of two digits is $L_z = 2(2S_f + 3S_d) = 18S_d$ and, thus, has one module width less than a digit encoded with the conventional code 2/5. When even numbers of digits have to be completed by a zero considerably more information can be packed on the same space.

[9] Because the start symbol in any case will be followed by other bar sequences the last bar has to be a white one.

124

Figure 6.8. Digits 124 in code 2/5 interleaved with check digit

While in code 2/5, the width (including quiet zone) is $b_{2/5}$ for n digits with $b_{2/5} = n \times 19S_d + 44S_d$ (with $n \in \{1, 2, 3...\}$) the code 2/5 interleaved for n digits has a fixed width of $b_{2/5i}$ of $b_{2/5i} = (n + (n\,mod\,2)) \times 9S_d + 29S_d$.

It has to be said, however, that the higher information density reduces the tolerance. The code 2/5 interleaved is self-checking but the error rate has increased considerably because the spaces (blanks) also include information.

Today more than 200 different ID-barcodes are known. [70].

6.2.4 Code 128

A typical and often used multi-width code is the code 128, which is based on four different widths as well as on three bars and three spaces per character and thus can represent the complete ASCII characters[10] from 0 to 127. The total of the different widths of the bars and spaces always corresponds to the width of 11 modules. The number of modules used to represent the bars is even and odd for the spaces. This allows for the early detection of errors already during the scan. The only exception is the stop character as the 106th character. It is the only character with 13 modules consisting of four bars and three spaces followed by a limit line two modules wide.

Like the code 2/5 the code 128 is a continuous barcode because the spaces are part of the code. In addition to the 128 ASCII characters, the representable character set of the code 128 consists of 100 digit pairs (from 00 to 99), four special characters, four control characters, three different start characters and one stop character. Owing to the three different character sets (codes A-C) the code 128 can represent more than 200 characters with 106 different barcodes.

[10] These are the digits, lower and upper cases, fullstop, comma, control symbols like tabs, backspace, enter... (ASCII = American Standard Code for Information Interchange, a worldwide used standard).

Table 6.2. Code 128

Value	Code A	Code B	Code C	Pattern	Barcode
0	SP	SP	00	212222	
1	!	!	01	222122	
2	”	”	02	222221	
3	#	#	03	121223	
4	$	$	04	121322	
5	%	%	05	131222	
6	&	&	06	122213	
7	’	’	07	122312	
8	(	(	08	132212	
9	)	)	09	221213	
10	*	*	10	221213	
11	+	+	11	231212	
12	,	,	12	112232	
13	-	-	13	122132	
14	.	.	14	122231	
15	/	/	15	113222	
16	0	0	16	123122	
17	1	1	17	123221	
18	2	2	18	223211	
19	3	3	19	221132	
20	4	4	20	221231	
21	5	5	21	213212	
22	6	6	22	223112	
23	7	7	23	312131	
24	8	8	24	311222	
25	9	9	25	321122	
26	:	:	26	321221	
27	;	;	27	312212	
28	<	<	28	322112	
29	=	=	29	322211	
30	>	>	30	212123	
31	?	?	31	212321	
32	§	§	32	232121	
33	A	A	33	111323	
34	B	B	34	131123	
35	C	C	35	131321	
36	D	D	36	112313	
37	E	E	37	132113	
38	F	F	38	132311	
39	G	G	39	211313	
40	H	H	40	231113	
41	I	I	41	231311	

Value	Code A	Code B	Code C	Pattern	Barcode
42	J	J	42	112133	
43	K	K	43	112331	
44	L	L	44	132131	
45	M	M	45	113123	
46	N	N	46	113321	
47	O	O	47	133121	
48	P	P	48	313121	
49	Q	Q	49	211331	
50	R	R	50	231131	
51	S	S	51	213113	
52	T	T	52	213311	
53	U	U	53	213131	
54	V	V	54	311123	
55	W	W	55	311321	
56	X	X	56	331121	
57	Y	Y	57	312113	
58	Z	Z	58	312311	
59	[	[	59	332111	
60	\	\	60	314111	
61	]	]	61	221411	
62	∧	∧	62	431111	
63	-	-	63	111224	
64	NUL	‘	64	111422	
65	SOH	a	65	121124	
66	STX	b	66	121421	
67	ETX	c	67	141122	
68	EOT	d	68	141221	
69	ENQ	e	69	112214	
70	ACK	f	70	112412	
71	BEL	g	71	122114	
72	BS	h	72	122411	
73	HT	i	73	142112	
74	LF	j	74	142211	
75	VT	k	75	241211	
76	FF	l	76	221114	
77	CR	m	77	413111	
78	SO	n	78	241112	
79	SI	o	79	134111	
80	DLE	p	80	111242	
81	DC1	q	81	121142	
82	DC2	r	82	121241	
83	DC3	s	83	114212	
84	DC4	t	84	124112	

Value	Code A	Code B	Code C	Pattern	Barcode
85	NAK	u	85	124211	
86	SYN	v	86	411212	
87	ETB	w	87	421112	
88	CAN	x	88	421211	
89	EM	y	89	212141	
90	SUB	z	90	214121	
91	ESC	{	91	412121	
92	FS	|	92	111143	
93	GS	}	93	131141	
94	RS	~	94	131141	
95	US	DEL	95	114113	
96	FNC3	FNC3	96	114311	
97	FNC2	FNC2	97	411113	
98	SHIFT	SHIFT	98	411311	
99	CODE C	CODE C	99	113141	
100	CODE B	FNC4	CODE B	114131	
101	FNC4	CODE A	CODE A	311141	
102	FNC1	FNC1	FNC1	411131	
103	Start CODE A			211412	
104	Start CODE B			211214	
105	Start CODE C			211232	
106	Stop			2331112	

A special feature of code 128 is that it can be arranged in lines so that active digits can be encoded which exceed the maximum reading width of a scanner.

6.2.5 Check digit calculation code 128

The code 128 also has a check digit p as the last sign before the stop character. For the character set z with n elements and $z = c_n \oplus c_{n-1} \oplus \ldots \oplus c_1$ this is calculated according to the formula:

$$p = \big(s + \sum_{i=1}^{n} w(c_i) \times i\big)\, mod\, 103$$

In this case w is an image which assigns a value to a character set according to Table 6.2, e.g.: $w(a) = 65$ or $w(b) = 66$.

As an example, the character sequence “sinus” should be encoded. Since this word consists only of lower case characters the character set code B is chosen. The character set is chosen by a character in the code, in this case the character with the value 104 (cf. Table 6.2). According to the formula above the check digit is built with $p = (s + 1 \times w(\text{“}s\text{”}) + 2 \times w(\text{“}i\text{”}) + 3 \times w(\text{“}n\text{”}) + 4 \times w(\text{“}u\text{”}) + 5 \times w(\text{“}s\text{”}))\, mod\, 103$.

The function w in Table 6.2 determines the values of the single alphanumerical characters $p = (104+1\times 83+2\times 73+3\times 78+4\times 85+5\times 83)\, mod\, 103$.

The check digit for the character sequence “sinus” is (when the start character is chosen for character set Code B) $p = 1322\, mod\, 103 = 86$ and has be added as character “v” (cf. Table 6.2) as stop character after the sequence “sinus”.

Start ⊕ “s” ⊕ “i” ⊕ “n” ⊕ “u” ⊕ “s” ⊕ “v” ⊕ Stop =

sinus

Figure 6.9. Character sequence “sinus” in Code 128 with check digit

This method for the calculation of a check digit applies for the code 128, all other symbologies use another method for this purpose.

6.2.6 The character sets of the code 128

The code 128 consists of the three character sets code A, code B and code C (*code in code*). The start character indicates whether the following code is a code A, B or C:

“Code A” “Code B” “Code C”

Figure 6.10. Start symbols for code A, code B and code C of the code 128

Code A contains, in addition to the upper case characters digits and punctuation characters, also the ASCII characters 0 to 31 which have a value of 64 to 95. Code B, on the other hand, contains in this range the lower case characters “a” to “z”[11]. Code C is completely different from codes A and B.

[11] Of course, umlauts can be represented when the ISO character set at the PC has been configured accordingly. Then the symbol “{” could be “ä” or the symbol “|” an “ö”. For compatibility reasons, however, special characters should be avoided in warehouse applications.

With the digits 0 to 99 it represents the digit pairs “00” to “99” so that two digits can be encoded by the bar sequence representing one character.

This is demonstrated by the following example where the digits “4465” are encoded with all three character sets.

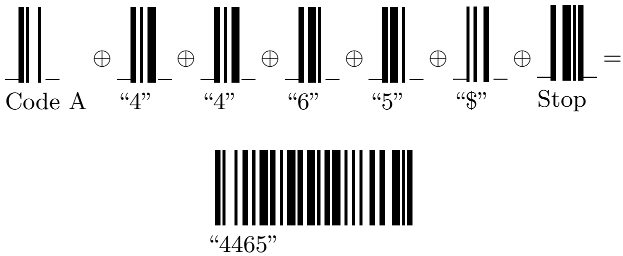


Figure 6.11. Digits “4465” with check digit in code 128 code A

In code A a start and stop symbol and a check digit are required in addition to the digits “4465”, i.e., a total of 7 bars. Since each character, except the stop character with a width of 13 modules, has a width of 11 modules the required space is $6 \times 11 + 13$ module widths. The stop character corresponds to a character with the value of 4, i.e., the character “$”.

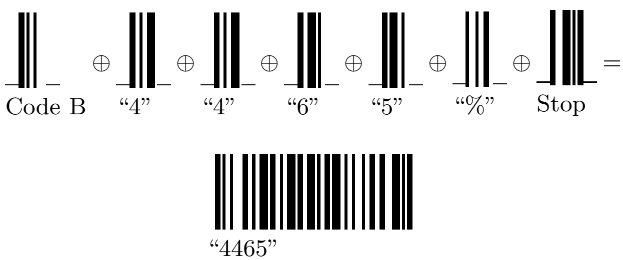


Figure 6.12. Numerical sequence “4465” with check digit in code 128 code B

In code B the same sequence “4465” starts with a different start character for code B. Since in code B the value (cf. Table 6.2) of the character is higher by a factor one than in code A the check digit is also higher by one, the character with the value 5, i.e., the symbol %.

When only digits are represented the density can be drastically increased by switching to code, which represents a pair of digits with one character, “44” as well as “65” are represented by one bar. A barcode in code C thus has 2 characters less or is by 2×11 module widths narrower than barcodes in code A or B. Regarding the active characters code C achieves a double density for numerical codes compared to codes A and B [49].

Code C ⊕ “44” ⊕ “65” ⊕ “73” ⊕ Stop =

“4465”

Figure 6.13. Numerical sequence “4465” with check digit in code 128 code C

Since a barcode in code C always requires an even number of digits a solution has to be found to represent odd numbers of digits. One possibility is to add a zero as first character.

6.2.7 Mixed character sets in code 128 and their optimization

In the code 128 the 200 characters of the three character sets can be mixed and combined freely. If lower and upper case characters and special characters are used in one sequence it can be switched between codes A, B and C. For this purpose, the code start character (103-105) is inserted. From that character on the character set of the other code applies until a stop character appears or the code is changed.

In addition to this, it can be switched from code A to code B and vice versa with the <SHIFT> symbol (character 98). The character following the <SHIFT> symbol (and only this) is represented in the character set of the other code. A switch to code C explicitly requires insertion of the character with the value 105 (code C) while a switch from code C to code A or B has also to be explicitly indicated.

A barcode can be optimized considerably by the clever use of the three different codes. However, to optimize and increase the information density it is necessary to analyze the characters to be encoded in advance. When more than four digits follow each other, the required space for the barcode can be reduced by switching to code C.

When a character sequence begins with two digits it does not matter if the barcode starts with code A or B or with code C. A start with code C, the representation of the digit pair and the switch to code A or code B needs three characters and thus 3×11 modules. The initial encoding of two digits with either code A or B has the same length.

The situation is different when the digit pair is not at the beginning of the character sequence to be encoded. In such a case a switch has to be made from code A or B into code C, the digit pair has to be represented, and then it has to be returned to code A or B. Three characters are required for this procedure. If, however, code A or B is maintained only two characters are needed to represent the two digits so that the required space is narrower by 11 modules widths. When four or more digits follow each other it is more efficient to switch to code C.

In case of an odd number of subsequent digits higher than four, it has to be determined when the code switch should take place. If the numerical sequence stands at the beginning of a character sequence the bar code should start with code C to minimize the required space. In case of an odd numerical sequence in the middle or at the end of a character sequence, code C should be used only after the first digit.

The switch from code A to code B has also to be studied more closely: A switch from code A to code B can be performed at any time but if it has to be initiated by a <SHIFT> symbol for a single character or for a longer character sequence depends on the individual case. If a single character can be represented by the character set of the other code the best solution is a <SHIFT>. In case of two digits, it does not matter whether a switch or double switch is chosen while for three or more characters a code switch is to be preferred.

If, on the other hand, two characters of another code are at the end of a character sequence the <SHIFT> symbol for these two characters would prolong the barcode by 11 modules compared to a switch because this would require no back-switch.

Generally, the choice of the optimal character set in code 128 is no trivial pursuit. Optimizations can well be achieved while the wrong switch may prolong the barcode drastically.

The characters <FNC3> and <FNC4> are included in the code 128 but reserved for special or future use. Here, FNC means “Function code”. The character <FNC2> induces the scanner to change into the multiline mode, to store the scans and to go to the beginning of the new sequence after the next scan. Since <FNC2> may appear several times this method allows for the reading of very long barcode sequences if supported by the scanner.

In combination with one of the three start characters the last character to be scanned <FNC1> introduces an EAN 128 code – quasi as a *twin start character*.

6.2.8 Code sizes, tolerances and reading distances

Certain sizes are recommended for each code for the print of module widths. Tolerances are also specified for each barcode which determine acceptable deviations of the bar widths [33]. These tolerances are of importance at least when barcode labels have to be printed with different methods and they may be a criterion in favour of or against a certain barcode.

The code 2/5 with a ratio of narrow bar to wide bar of 3:1 and a ratio of narrow bar to separator of 2:1 has a tolerance of 20%. This is very high and due to the fact that the spaces between the bars do not hold any information. Thus, the code 2/5 is optimal for a simple print on matrix or ink-jet printers.

The code 2/5 interleaved has a considerably lower tolerance, which is based on the ratio between the narrow and wide element v (between 2:1 and 3:1) and the chosen module width m_b. Thus, the following applies to the tolerance t of the code 2/5 interleaved: $t = \pm((18v - 21)/80)m_b$ [49].

In case of a ratio between the narrow and the wide element $v = 3$ and a chosen module width $m_b = 0,33$mm the tolerance (rounded) is $t = \pm((18 \times 3 - 21)/80) \times 0,33mm= 0,136$mm. Thus, in the code 2/5 interleaved and at a ratio between narrow and wide elements of $v = 3 : 1$ the tolerance t must not exceed one third of the module width. If the ratio valency sinks the tolerance is automatically reduced as well. In case of a ratio of $v = 2$ it is only $0,062$ mm.

Due to the low tolerance of the code 2/5 interleaved compared to the code 2/5, this code requires a higher printing quality and thus a more expensive technology.

The acceptable tolerances are specified for each barcode symbology. Three different tolerances b, e and p are given for code 128 which depend on the module width m_b:

- $b = \pm$ 0,33m_b the tolerance of bars and spaces,
- $e = \pm$ 0,2m_b the tolerance of bar edges within a character and
- $p = \pm$ 0,2m_b the tolerance between the first bar of a character to the first bar of the following character [103].

This low tolerance is understandable because the code 128 is a multi-width code with information-bearing spaces.

There are different recommendations for the module widths of the single codes. Values from 0.3 mm to 0.35 mm can be found, while the general enlargement factor 1.0 of the barcode is 0.33 mm [49, 70, 103]. An enlargement factor larger than 1.0 (about 1.35 or 1.5) improves the readability but requires more space. The reading distance as maximum scanner distance which still allows for a faultless scan depends on the scanner and the chosen enlargement factor of the barcode. The reading distance is given by the manufacturer of the barcode scanner either as maximum distance (e.g., 100 mm) or as scanning range (e.g., 100 mm - 500 mm).

In practice, many errors[12] are due to the non-observance of reading distances. Depending on the scanner the distance may be 0 mm (touch reader) (cf. section 6.5.2) or several meters (stationary industrial scanner or camera system, (cf. section 6.5.3). Light also plays an important role with regard to the reading distance since not all barcode scanner have a light source of their own (e.g., cameras). Furthermore, the correct print depends on the pitch and the screw of the scanner with regard to the barcode as well as the print contrast signal.

[12] Here, errors mostly are the "NoReads", i.e., unread scans.

6.3 Printing method and quality

To use the barcode for the artificial encoding of an object, it has to be generated and applied prior to the actual identification. The most optimal method for the respective application has to be chosen from the variety of printing and labelling techniques.

6.3.1 Labelling techniques

In principle, all printing techniques are suitable which print barcodes on the desired surface with an adequate quality and precision. There are two different labelling methods, the direct and the indirect labelling:

Table 6.4. Labelling techniques

Direct labelling	Indirect labelling
– Direct ink-jet print – Direct laser labelling – Engraving – Pins	– Matrix print – Ink-jet print – Laser print – Thermo transfer print – Direct thermo print – Photosetting / Offset print – Screen printing

In case of a direct labelling the barcode is directly printed on the object without a separate carrier. Ink-jet printers are often used to label cardboard boxes. Other materials (plastic, glass, metal) are best labelled with laser printers or by mechanical engraving. The quality of the prints largely depends on the materials used and is reliable only under unchanged conditions. Despite of their poor quality direct labels have some advantages:

– Cost savings (no labels or applicators)
– Code generation can easily be automated
– Code is directly and inseparably applied to the object
– A variable code and encoding in plain writing is possible
– Ink labels can be removed from suitable surfaces and the objects can be relabelled in automatic cycles

The second group of labelling techniques is used more often above all in logistic systems, i.e., the indirect labelling. Here, at first a label or code applicator is printed, which is then attached to the object. This may also be done later.

6.3.2 Quality requirements

The quality of the barcode print has to ensure the quick, unproblematic and, above all, faultless scan of the code. Especially in highly automated warehouses each unidentified object initiates a system error and, thus, high costs. The main reason for wrong reads is the poor quality of the *auto-ID label*. Because of the high conveying speed in automatic warehouses it is vital to achieve a high initial scan rate, i.e., the code has to be identified already by the first scan line. Important quality criteria for the printing of barcodes are [48, 51, 68, 100]:

- Contrast between bright and dark
- Dimensional accuracy of the print
- Contour sharpness
- Congruence of the black surfaces
- Resolution (for very small codes)
- UV resistance
- Scratch and smear resistance
- Water and solvent proved

Above all the contrast largely depends on the used source of light. A red barcode on a white background, for example, cannot be read under a red laser light (cf. section 6.5.1). The most tried and tested combination is black on white [48].

6.3.3 Selection of the printing technique

The quality of the barcode is of great importance for the choice of a printing technique although not the sole criterion. In case of doubt the decision in favour of or against a certain technique is based on the purchasing and operating costs. The following aspects have to be cleared before choosing a printing technique:

- Availability of the data to be encoded in the long run or a little bit in advance
- Changeability of the data contents
- Direct or indirect labelling (print on the object (book, packaging, workpiece) or on a label)
- Required printing speed
- Media handled by the printer (e.g., form, thickness and structure of the label)
- Barcode to be used
- Space available for the barcode
- Advanced quantitative determination of the needed labels
- Purchasing costs

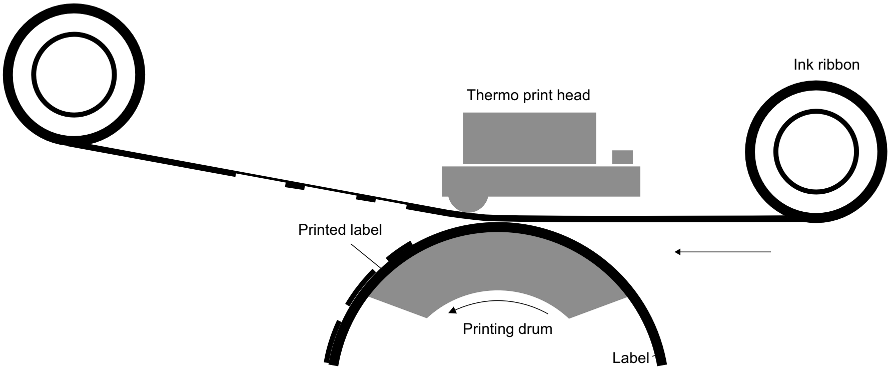


Figure 6.14. Functioning of a thermal direct printer

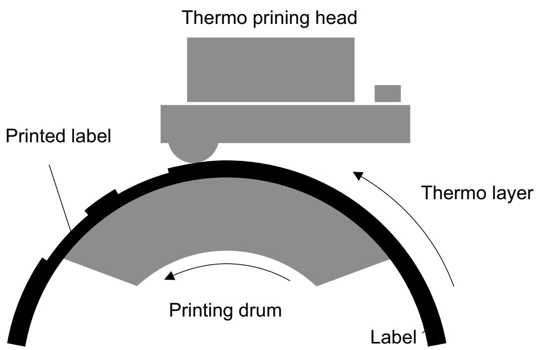


Figure 6.15. Functioning of a thermal direct printer

– Estimated costs for consumables such as toner, ink or labels as well as for maintenance and repair of the equipment.

In any case, special attention has to be given to the ecological acceptability of the used printers and materials. In some cases, this aspect may be decisive when label printing is outsourced to other companies. During operation labels can be printed with different techniques: While matrix, ink-jet and laser printers are widely used in offices industry mainly works with thermal direct and above all thermal transfer printers. Here, the label is guided along a print head where small ceramic elements can selectively be heated. At a thermal direct print the heat initiates a chemical reaction on the labels which causes colored spots on the label. This method requires thermo-sensitive labels.

In case of a thermal transfer print (cf. Fig. 6.14) a wax or resin-coated ink ribbon runs between the print head and the label. The heat melts the ink particles, which are then transferred to the label. The print need not be fixed and is immediately smear resistant.

Thermal direct printers are simple and highly compact systems (cf. Fig. 6.15). Their advantage is the small quantity of consumables (special paper) so that maintenance and operating costs can be kept low. The printers are easy to operate, work at a medium speed and provide a good quality. Labels printed by a thermal direct printer are not too suitable for a long-term use because the chemical reaction in the label may be reversed so that the label would become sensitive to heat and UV rays.

Table 6.5. Characteristics of laser and thermal transfer printers

Characteristics	Thermal transfer	Laser
Print quality	++	++
Smear and scratch resistant	++	+
UV resistance	++	++
Speed	+	++
Robustness	++	o
Purchasing costs	−−	o
Operating costs at		
high share of black	o	−
low share of black	o	+

Thermal transfer printers are more sensitive because of a more elaborate mechanics and, thus, generally cause higher operating costs. The printing costs remain constant since a fixed length of ink ribbon is used for one label. The costs for a 20% black print; however, (cf.Table 6.5) are lower than for a laser print [68]. The prints are of high quality, smear, heat and UV resistant as well as water, alcohol and fat proved.

The decision whether conventional laser-ink-jet or matrix printers should be used instead of the relatively expensive thermal transfer printers can only be made after *all* parameters have been taken into consideration. Special attention has to be given to the requirements on the printing quality and the durability of the print. If these requirements are high and if the print should be resistant to all external impacts, a thermal transfer printer is a good solution [68] while laser and matrix printers are good solutions for office and industrial use.

6.4 Semantics in the code: EAN 128

While the code 128 can be used to encode data on a barcode level the code EAN 128 is used for a cross-company communication. The barcode symbology of the code 128 defines all technical basic characteristics of a code. However, the organizational handling of the barcode is not determined so that the definition of code 128 can be seen as *basic layer* of an encoding technique which does not yet provide any rules regarding the future practical use. These are provided by the code EAN 128.

Any warehouse operator is able to define objects which are necessary for the warehouse operation, such as articles and their packaging as well as load supports and storage bins for the logistic processes and to label them by means of barcodes. Such labelling is indispensable for automatic identification. As long as the objects do not leave the in-house cycle, i.e., the company premises, it is sufficient to label them with an internal code. This is best done by a systematic approach, such as the redecoding or uniqueness of the numbers.

However, for cross-company communication, a decoding method has to be agreed upon to allow for a "common language". Such an agreement is very complex and has to be repeated every time a partner wants to join the commercial or communication group. Another possibility is to use an industry-wide standard such as the code EAN 128. In 1977 in Germany the CCG, "Centrale für Coorganisation GmbH — Gesellschaft zur Rationalisierung des Informationsaustausches zwischen Handel und Industrie" was founded. The CCG represents the "International Article Numbering Association" (an umbrella organisation located in Brussels). Both organizations focus on the further development of the EAN systems which

> "in the meantime has become a worldwide standard for identification techniques, or even better, the only standard for really international applications" [12].

The EAN systems are based on three numbering or encoding systems, the GLN (global location numbering), the EAN (international article numbering)[13] and on the SSCC (aerial shipping container code). These three encoding systems are included in the code EAN 128 which, on the other hand, can also be represented in the barcode symbology of the code 128.

6.4.1 Global location numbering (GLN)

The GLN serves as physical address of companies and sections or departments. It is compatible to the GLN concept of EAN and applies worldwide. A GLN is assigned just once and can be tracked down precisely.

The GLN is divided into type 1 and type 2, which are compatible. Global location numbers of type 1 are used for the clear identification of an organization, but not for the numbering of articles or unit loads. A type 1 GLN consists of 2 digits and one check digit. In Germany this number is assigned consecutively by the EAN commission, the CCG.

A GLN type 2 also consists of 13 digits (twelve digits and one check digit). Here, the EAN authorities assign just a seven-digit basic number while the

[13] Originally, the term EAN stood for European article numbering.

other five digits and the check digit are freely determined by the GLN user and should not be used otherwise.

GLNs of this type are used to build continuous location numberings for own sections such as production lines, warehouses, departments or other locations and as article numbers. Up to 100 000 company sections or locations can be defined with these digits.

Table 6.6. Examples of country codes: Germany and its neighbours

Numerical sequence	Appropriate organization
30 to 37	GENCODE EAN France
400 to 440	CCG Germany
54	ICODIF / EAN Belgium and Luxembourg
57	EAN Denmark
590	Poland
76	EAN Switzerland
859	EAN Czech Republic
87	EAN Netherlands
90 to 91	EAN Austria

The first two or three digits (cf. Table 6.6) of the GLN are the country code. This code should not indicate the country of origin of the product or the location of the company but the country where the location number was assigned. The general terms and conditions of EAN Austria stipulate, for example: "Eligible are all companies which participate in the goods and services traffic in Austria."[14] A company with headquarters outside the country may have a GLN in this country but also in all other countries worldwide. The GLN is a company code, but not a code of origin. Table 6.6 shows the codes of different countries.

A company has been assigned, for example, the GLN 40 12345. It has offices in Berlin, a warehouse in Bochum with 4800 storage bins and a production site in Bremen. With the GLN the company can determine location numbers for each of its three section. Since these numbers have to have five digits corresponding zip codes may be used here, i.e., 10713 for the offices in Berlin, 44791 for the warehouse in Bochum and 28777 for the production site in Bremen.

Together with the corresponding check digit the valid thirteen digit GLNs thus would be 4012345107135, 4012345447910 and 4012345287776. It would be wiser, however, to assign number ranges to the single locations, e.g., the range from 40000 to 49999 for the warehouse in Bochum. Finally this range should be refined, e.g., by assigning the number 40000 to 44799 to the storage

[14] cf.: http://www.ean.co.at

bins, the number 44800 to the goods receipt and 44801 to the master office, etc. If a company has a GLN type 2, it may built systems of its own and assign the five location numbers at discretion.

6.4.2 International article number (EAN)

The EAN guarantees for the clear international identification of single articles. Consisting of machine readable barcodes, it sets the basis for automatic scans at the point of sale (POS).

Each company with an international location numbering type 2 is able to generate EANs of its own. For this purpose, the first seven digits of the GLN (type 2) build the basic number. Like in sub-locations the next five places can be filled with own digits which correspond to 100 000 different articles. The last place is reserved for the check digit. The result is a correct thirteen digit EAN. Should these 100 000 options not be enough, i.e., if a company offers more than 100 000 different articles another GLN can be requested.

The GLN and the EAN are very much alike and can easily be mixed up. According to the CCG, however, it can be derived from the context if a code is a GLN or an EAN so that a confusion is very unlikely in practice.

One speciality is the EAN 8, an eight-digit EAN code which was developed above all for small-volume goods and articles. Since these very short EANs have a limited number range the existence of the related article has to be proved to the corresponding EAN authority.

For example, a company has been assigned the GLN 40 12345 and has an assortment of five articles, it can number these articles itself, e.g. 00001 for the first article, 00002 for the second, etc. Together with the check digit the EANs would be 4012345000016, 4012345000023, etc. The purpose of a classification of articles into groups and a subsequent segmentation will not be discussed here because experience has shown that so-called *descriptive number ranges* cause considerable problems at least when the system boundaries are exceeded.

With regard to licence agreements, each company participating in the EAN system is obliged to ensure a clear numbering and to block numbers for a certain period of time if the related articles are no longer available. Thus it is ensured that, for example, in a supermarket the master data totally correspond to the articles.

6.4.3 Serial shipping container code (SSCC)

The basic number of a GLN serves as basis for the generation of SSCC. The term *shipping* describes a logistic unit (cf section 2.1.2), i.e., physically connected units, e.g., packages like pallets or cardboard boxes which cannot be separated.

The recipient of a package, who may also be a sender (forwarder), can continue to use the SSCC of the container as long as he does not break the unit (the package). He may notify the next recipient about the shipment just by mentioning the received SSCC.

The SSCC is an 18-digit numerical sequence with the following sections: The first digit is the reserve digit, generally the digit 3. In the USA and Canada the UCC[15] has used this reserve digit of the SSCC to describe the type of shipping container (such as cardboard box, pallet, crate or parcel). In a SSCC the digit 3 stands for an "undefined shipping container". However, since the adjustment of both systems EAN and UCC on 1 January 2001, the digits 0 to 9 can freely be used as reserve digits. The reserve digit is directly followed by the basic GLN number of the sender so that the shipping container can clearly be assigned to the shipper similar to the EAN[16].

Then follows a continuous nine-digit number that is freely assigned by the user and, if necessary, has to start with an adequate number of zeros. According to the CCG this number should not be reused for a certain period of time — one year is proposed — to ensure the uniqueness of the SSCCs. If the nine digits are not enough, i.e., should more than one billion shipping units be built within this period, another GLN can be requested at the EAN authority to generate further SSCCs. The last digit in the SSCC is the check digit which is built according to the code 2/5 from the right side with the weighting factors 3, 1, 3, 1, ... and the reloaded module operations[17].

For example, a company has been assigned the GLN 40 12345, the continuous number of the shipping container is 987654321 and the check digit calculated according to the formula of code 2/5 is 5, together with the reserve digit 3 the SSCC is: 3 40 12345 987654321 5.

Provided that the SSCC is correct, unique and well documented the shipping container can be identified worldwide. Owing to the GLN basic number and the standardization of EAN and UCC each shipping container can be directly assigned to a company. This sets the basis for an open and worldwide tracking and tracing of consignments, which in the past has been an additional service but nowadays is an indispensable feature and simplified considerably by SSCC.

6.4.4 Characteristics of the code EAN 128

The code EAN 128 can be generated by directly including <FNC1> in the code 128 after the start symbol.

The code EAN 128 not only carries different data like the SSCC, the GLN or the EAN, but also a number of other types of information. This information is combined by a number of qualifying *data identifiers* and can be represented in the form of barcodes. A fixed format of the data identifier in relation to

[15] United Code Council, an organization pursuing the same aims as the EAN.
[16] Since 1st January 2001 the CCG also assigns 8-digit and 9-digit basic numbers but an SSCC label must not exceed 18 digits.
[17] The check digits for GLN and EAN are built by just the same method.

its contents ensures that different data identifiers and the related contents can be written one after the other. There are two-, three- and four-digit data identifiers as well as data field contents of fixed and variable lengths. In addition to numerical contents, some data fields may also contain alpha-numerical data. When different combined data fields are represented by the same barcode, the fields with a fixed length should be at the beginning while data fields with a variable length which have to be separated by a <FNC1> are placed at the end.

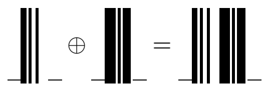


Figure 6.16. Symbols of the code B and <FNC1>, the double start character

The data identifiers are described, among other places, in [12], where it is defined whether a data field is variable or fixed, how many digits are provided for the identifier, and how many for the contents. Since all data identifiers are numerical, it makes sense to do without alpha-numerical symbols as far as possible and develop adequate codes for these fields instead. With its completely numerical character set, the code C of the code 128 can be used throughout; this compression saves up to 50% of the space[18] (cf. section 6.2.7).

Table 6.7. Examples of data identifiers (DI) and contents

DI	Data contents	Length	Speciality
00	Number of the shipping container	2+18	
01	EAN of the sales unit	2+14	
02	EAN of the included unit	2+14	
13	Packaging date yymmdd	2+6	tt $\in$ {00..31}
15	Minimum durability yymmdd	2+6	tt $\in$ {00..31}
37	Number of included units	2+n	n $\leq$ 8 (fnc1 at the end)
410	Recipient's GLP	3+13	
412	Supplier's GLP	3+13	
99	Bilaterally agreed texts	2+n	n $\leq$ 30 (fnc1 at the end)

Table 6.7 shows an example of some data identifiers and contents as well as the space they require. The date in the identifiers 13 and 15 has to be followed by a six-digit information. If no day is given 00 has to be entered.

[18] Except the start, stop and <FNC1> symbol as well as the related check digit which cannot be compressed.

The data identifier 99 is an example of a variable length and the following information may consist of up to 30 digits (in code C up to 60 digits). If the identifier 99 is followed by another data identifier the separator <FNC1> has to be put at the end of the field to mark the beginning of the next field. Therefore, the <FNC1> symbol must never be used as a text sign.

The data identifier for the SSCC is 00 (cf. Table 6.7) and together with the 18-digit SSCC it requires space for 20 digits. If, for example, a barcode of the type EAN 128 should be generated with the SSCC 3 40 12345 987654321 5 this looks as follows:

Code C ⊕ FNC1 ⊕ 00 ⊕ 34 ⊕ 01 ⊕ 23 ⊕ 45 ⊕

⊕ 98 ⊕ 76 ⊕ 54 ⊕ 32 ⊕ 15 ⊕ 58 ⊕ Stop =

00 3 40 12345 987654321 5

Figure 6.17. Correctly generated NVE in the form of a barcode

The SSCC built according to Fig. 6.17 can be used and printed on a transport label for direct application. Transport labels should just bear the SSCC and no further information.

The reason for this becomes evident when looking at mixed pallets: The data identifier 02 defines the EAN of the included sales unit and the data identifier 37 the quantity of the included units[19] as to determine the single contents and quantities. Since, however, there is no regulated order of information about sales units and quantities (e.g., a meta data identifier which includes the identifiers 02 and 37) this may lead to some misunderstandings. A mixed pallet, for example, requires twice the identifier for the EAN and the quantity. In the age of electronic communication the contents and the quantity related to a SSCC can be sent by fax or *EDI* (electronic data interchange). New promising methods like, for example, *AS2* (Applicability Statement 2) by IBM or *BPEL4WS* (Business Process Execution Language for Web Services) by IBM, Microsoft and BEA are used in first applications.

[19] According to Table 6.7 the identifier 37 has a variable length. If it does not stand at the end of a barcode its contents has to be separated from the next identifier with a <FNC1> separator.

Code A ⊕ FNC1 ⊕ 0 ⊕ 0 ⊕ 3 ⊕ 4 ⊕ 0 ⊕ 1 ⊕

⊕ 2 ⊕ 3 ⊕ 4 ⊕ 9 ⊕ 8 ⊕ 7 ⊕ 6 ⊕ 5 ⊕

⊕ 4 ⊕ 3 ⊕ 2 ⊕ 1 ⊕ 5 ⊕ 'Z' ⊕ Stop =

00 3 40 12345 987654321 5

Figure 6.18. NVE in the unpacked code A

Fig. 6.17 shows a SSCC generated from code C of code 128. By combining two digits the numerical sequence 98 76 54 32 15 can be represented with a module width of $14 \times 11 + 2$

00 3 40 12345 987654321 5

00 3 40 12345 987654321 5

Figure 6.19. Comparison of the lengths of code A and code C with same contents

Of course, the SSCC of the numerical sequence above can be represented by code A or B of code 128, as is shown in Fig. 6.18. This requires $23 \times 11 + 2$ modules and thus 99 module widths more than the most space saving representation in code C. A numerical sequence with a module width of 0,33 mm in code A has a width of 84,15 mm and in code C of 51,48 mm (cf. Fig. 6.19). Common and widely used barcode scanners in the form of a handheld touch reader with a scan width of 60 mm up to max. 80 mm would not be able to read the unpacked form of code A or B.

6.5 Scanner technology, devices, interfaces

The choice of a suitable barcode symbology for the use and generation of the codes and the application on a object is just *one* problem to be solved. The scanning, identification and processing of the barcode is another which will be studied in detail in this section. When a barcode is scanned the optical image in the form of light and dark bars of different widths has to be scanned, digitized, identified and processed into mechanically readable data flows. All this is performed by a barcode scanner.

6.5.1 Barcode scanner

A scanner and a decoder are the basic parts of a reader. The scanner receives the reflected signals of the bars and spaces of a barcode with a single photo diode, an array of photo diodes, a photo transistor and a *CCD* (charge-coupled-device).

For this purpose, the barcode has to be sufficiently lighted and the patterns have to have a sufficient print contrast signal. Older devices have to have a print contrast signal of a least 70% to ensure a correct scan, newer devices can be operated with half the signal. The print contrast signal d_k can be determined based on the reflection of the bars r_s and the spaces r_l with the formula: $d_k = 100\% \times ((r_l - r_s)/r_l)$. If, for example, the background of a barcode has a reflection of 80% and the bar of 42% this results in a print contrast signal of $d_k = 100\% \times ((80\% - 42\%)/80\%) = 47.5\%$. If a barcode on such a medium should be readable for older devices the reflection of the darker bar r_s has to be $(70\% = 100\% \times ((80\% - r_l)/80\%) \Leftrightarrow r_l = 24\%)$ and should be reduced to 24 percent. Generally, this is achieved by a darker colour.

The light reflected by the barcode is digitalized by the scanner with a A/D converter and transmitted to the decoder. In the past, scanner and decoder were separate devices, but now they are built as a unit in the same casing. The decoder decodes the transmitted digital data into the ASCII code and provides the decoded numerical or character sequences.

Prior to studying the different barcode readers, they are classified into portable and stationary devices.

6.5.2 Handheld scanners

The first barcode readers on the market were barcode pens that had to be moved directly above a barcode at a continuous speed. These simple structured devices are still in use sometimes. A light ray bundled by a lens or a laser diode is sent to a point at the light-sensitive end of the scanner (cf. Fig. 6.20).

This end is moved along the barcode, which reflects the light more (light surface) or less (dark surface). A photo diode converts the reflected light into

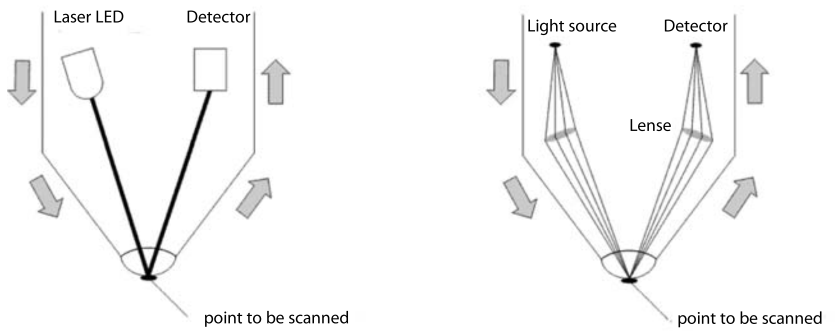


Figure 6.20. Scanner

a voltage. By smoothing the voltage edge (Schmitt trigger principle) and levelling the time a digital signal can be generated and thus the bars, spaces and their respective length can be interpreted.

More efficient than code pens are CCD handheld scanners or scanners with a mobile ray (cf. Fig. 6.21). The CCD handheld scanner, which are primarily in the form of touch readers, contains an array of microscopic photo diodes which can scan the barcode as a complete image. The barcode is lighted by an LED line which emits red or infrared light. These user-friendly, light and robust devices are ideal for order-picking terminals, cash points and manual work places; however, the scanning range is limited by the CCD array is of disadvantage.

The portable scanner with a mobile ray sends a light signal, which is generated by a semi-conductor laser diode, via a mirror oscillating at 30 or 50 Hertz or a rotating polygon mirror onto the barcode. The reflection of the lighted areas is recorded by photo detectors. This device is able to read longer scan lines and thus longer barcodes but due to the required mechanics it is larger. More comfortable devices dispose of processors with multi-line mini-displays and block keyboards. Some of them offer terminal functions or the possibility to directly preprocess the data.

Handheld scanners generally have an RS232 interface. Some smaller handheld CCD scanners can even be plugged between the keyboard and the computer. The barcode scanner then works transparently i.e., like a computer keyboard.

6.5.3 Stationary scanners

In addition to handheld devices, stationary barcode readers are used . These look and function like the handheld scanners. The only exception are the camera systems that do not scan the code line by line but process the complete image. A commonly used version is the POS scanner (point of sale), which is integrated in shop counters and was developed to scan barcodes at consumer goods.

Stationary scanners are used at conveyor lines for the identification of piece goods where conditions remain relatively unchanged to allow for an automatic scan or where a manual scan is impossible for throughput reasons. In front of shunts they collect data for transport decisions, behind the shunt for control purposes.

A significant difference between the handheld scanners and stationary barcode scanners is the data transfer method. While most of the handheld scanners transmit the data via the keyboard port or RS232, many stationary industrial scanners are connected to a RS48 interface via various field bus systems. This allows for a data transmission over considerably longer distances (cf. Chapter 5).

6.6 2D-Codes

In addition to ID-codes, simple linear (horizontal) barcodes or optical codes are used to increase the information contents. These codes also use the 2nd

Figure 6.21. Handheld scanner [INTERMEC TECHNOLOGIES]

dimension, the "vertical" component for the transfer of information. In the most simple case common barcodes are put one above the other, i.e., in multiple lines. More complex methods reduce the bar to a dot and are thus called *matrix* or *dot codes*.

The 1D-barcode is a reference to an information while the actual information is transmitted by a 2D symbology with a higher information density.

6.6.1 Stacked barcodes

In 1987 the company Intermec was the first to develop the stacked barcode of the code 49 to be used in astronautics. With this code a maximum of 81 digits or 49 alpha-numerical signs can be transmitted which are encoded in up to eight lines. The basic idea was to better utilize the barcode space for data and to minimize the redundancy of prolonged bars by a multi-line representation of the codes. This clearly reduces the height of the single barcodes.

Figure 6.22. Example of a code 49

One year later German engineers started to develop another stacked barcode, the Codablock barcode. In a Codablock a barcode line is continued until the line is complete; then it is wrapped. In Codablock F, which is based on the code 128, up to 44 lines with four to 62 characters each can be encoded, that is a maximum total capacity of 2728 characters. Nowadays, Codablock is widely used in health care, e.g., to mark blood bottles.

Figure 6.23. Example of a Codablock barcode

A widely used stacked barcode is the PDF 417 barcode (Portable Data File). Each single character is encoded into a codeword with a width of 17 modules which consists of four differently wide bars and spaces. A maximum of 90 lines with 2700 digits or 1850 ASCII characters can be stored.

Figure 6.24. Example of a PDF 417

Of course, it requires another technology to read 2D codes, i.e., camera systems and *fan scanners* in addition to line scanners.

6.6.2 Matrix codes

Only few of the many matrix codes are commonly used, such as:

- Aztec
- QR Code
- MaxiCode
- Data Matrix Code
- Dot Code A

In case of the *Aztec code* several nested squares in the middle of this square code serve as *search element*[20]. 12 to 3000 characters can be encoded and owing to an error correction the code can even be read when up to 25% of the data surface are damaged.

Figure 6.25. Example of an Aztec code

The *QR code* (quick response code) consists of nested squares situated in three corners which serve as a search element. This code is also square and owing to an error correction the data can be reconstructed when up to 30% of the data surface are damaged. The code was developed in Japan and can store up to 1810 Japanese Kanji/Kana characters, 4296 ASCII characters or about 7000 digits.

For the shipment of parcels the MaxiCode is used where 93 ASCII characters or 138 digits are encoded on a fixed surface of one square inch (25,4 mm×25,4 mm). The Max Code has a search pattern consisting of three

[20] The search element serves as reference point for the image processing software.

Figure 6.26. Example of a QR code

concentric circles around which 866 hexagons are arranged in 33 lines which are filled in black or white. Owing to the MaxiCode error correction data damaged up to 25 percent can be reconstructed.

Next, there is the Data Matrix Code, version ECC 200 , which was developed in the USA in the late 1980s. Up to 1558 ASCII characters or 3116 digits can be encoded on a square surface of a variable size.

Figure 6.27. Example of a Data Matrix Code

The search element of the Data Matrix Code is a horizontal bar at the lower end and a vertical at the left side. The alternating black and white pattern at the left and the upper side defines the size of the symbol and thus of the data contents. The Data Matrix Code is mainly used in the electronic industry to label conductor boards and and chips as well as in the automotive industry and as digital postage stamp.

Figure 6.28. Example of a dot code A

The *Dot Code A* consists of a square of 6×6 up to 12×12 dots. Up to 42 billion objects can be encoded with this code which is used, for example, to identify laboratory tubes, to label linen in laundries, but also to process aluminium slugs with stamped dot codes [50].

Last but not least, there are the so-called *3D codes*, which are based on the 2D codes and use colored components as a third dimension.

6.7 Radio frequency identification

The introduction of automatic identification technologies supported or even allowed for the tracking and tracing in many logistic applications. Of all available technologies barcode systems are the most commonly used, owing to their simple structure, good standards and low costs. But the *radio frequency identification* (RFID) is becoming more and more important.

This section deals with the RFID technology and its use in logistics. In fields where barcodes are used RFID systems offer additional features and open up a variety of new fields of application. This study from the technical and organizational point of view is followed by an comparison between barcode and RFID.

6.7.1 Functioning and technical structure

RFID is a technique for the contact-free recording and transmission of binary encoded information by means of electromagnetic waves. Information are transmitted between an information tag[21] applied to the goods and a mobile or stationary scanner. RFID systems are divided into two main function groups:

- *Transponders*[22] as information carriers on the goods to be identified. The information stored on the transponders are just readable or rewritable, according to the type.
- A device for the recording of transponder information. Depending on the transponder type either readers or writers/readers can be used.

Since the data are transmitted via radio transmission, no intervisibility is required between the transponder and the scanner during the transmission.

All types of transponders are equipped with an antenna for the data transmission, a transmitter to couple antenna signals[23] as well as a microcontroller (μC) to control the data in the memory. Controller and memory often are integrated in one chip (cf. Fig. 6.29).
The active data received by a write/read device can be transmitted via a serial interface (RS232, RS485) to a computer for further processing.

RFID systems are available in a variety of forms (cf. Fig. 6.30). Basic differences exist with regard to the power supply, the memory technology and the frequency ranges used for data transmission.

[21] The most commonly used term for RFID carriers is *transponder*, but terms like *mobile data store* (MDS) and *tag* are also used.
[22] The term is combined of **trans**mit and res**pond**.
[23] Demodulation of the data received by the transponder or modulation of the sent data.

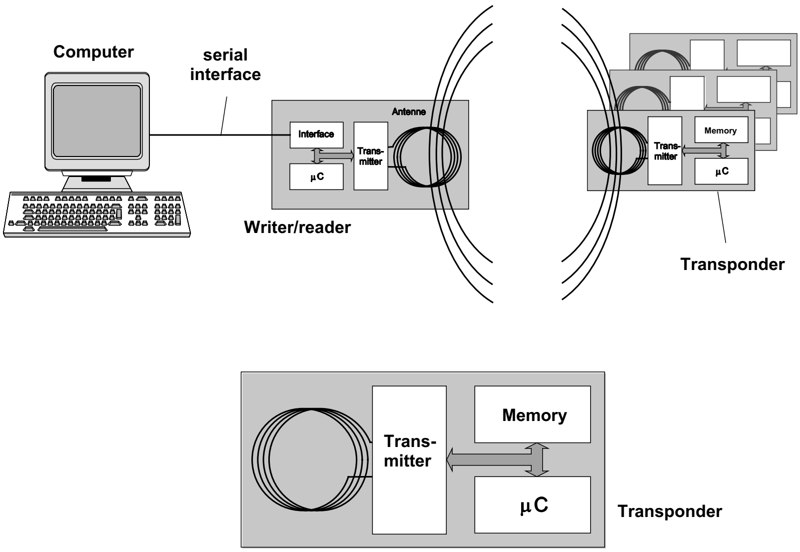


Figure 6.29. Function groups of RFID systems

Power supply The power supply mainly affects the size, lifetime and application range of a transponder. Transponders, thus, are divided into two basic categories:

Passive transponders are powered by a magnetic field generated by the scanner. As long as a transponder is within the range of the scanner the externally radiated alternating field generates a voltage in the antenna coil. The power is transmitted according to the transformer principle but only over short distances.

Active transponders are equipped with a battery (flat or coin cell) for the chip and can independently initiate a data transmission. Owing to the internal power supply, the performance is large enough also for long-distance data transmission.

In a certain kind of active transponders the battery is used only to buffer the data memory. The required power for the transmission and reception is — like in passive systems - generated by the alternating field.

Memory types The data memories of transponders have a capacity of one bit[24] up to 128 Kb. *Read-only* memories contain fixed actual data already

[24] 1-Bit transponders for the electrical article securing (EAS), e.g., in the textile industry.

Figure not drawn to scale

Figure 6.30. Forms of transponders: Device with external antenna (A), foil transponder as adhesive label (B), glass transponder for animal injectates (C), container transponder (D)

determined by the manufacturer. The buffering of data needs no electricity because a ROM[25] is used and the access speed is relatively low. Therefore, they are used for the storage of small data quantities (e.g., article numbers) with a typical capacity of some hundred bits. *Write-once* memories contain similar technical data and are used for the same purposes but can be written once by the user (OTP-ROM[26]). In *Read-/Write* memories RAMs[27] are used which allow for a quicker access and have a larger capacity but require a battery buffer.

Data transmission, frequency ranges and coverage The design of a transponder antenna depends on whether it is a wire coil, a print coil[28] or a dipole (for microwave transmission). If a transponder is located within a transmission field, the chip is activated and in case of read-only transponders the stored active data are transmitted.

In passive systems the data are transmitted in two phases: During the first phase the field energy is used to charge a memory condensator, which in the second phase activates the chip and thus the data transmission. This procedure is continued until the transponder is no longer activated by the trans-

[25] Read-Only-Memory
[26] One-Time-Programmable ROM
[27] Random-Access-Memory
[28] In Smart Labels coil and chip are laminated on a foil.

Table 6.8. Common RFID frequency ranges (Europe)

Frequency range	Transponder type	Coverage
125 - 135 kHz (LF)	Passive	Some cm up to 1 m
13,56 MHz (HF)	Passive	Up to 1 m
433 - 868 MHz (UHF)	Active	Several m
2,45 GHz (Microwave)	Active	Up to several hundred m

mission field. In case of active systems the first phase is omitted so that the response time may be shorter. Rewritable transponders require a protocol-based transmission where either the reading or writing of the transponder is initiated by the transmitter by means of a corresponding telegram. . Principally, higher transmission frequencies allow for a larger coverage and thus for higher transmission speeds (Table 6.8).
The achievable transmission rates and coverage depend on the physical environmental conditions[29] and are furthermore defined by corresponding standards and regulations.

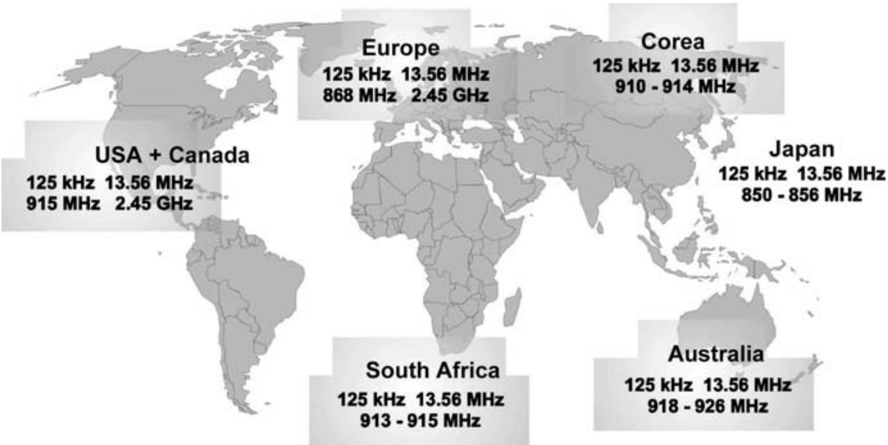
Allowed frequency ranges and transmission rates for transponders in Germany and Europe are stipulated by corresponding regulations[30] as not to interfere with other radio services (mobile telephony, broadcast) [97].

In the USA, South Africa and Asia further frequency ranges are allowed so that passive transponders and microwave transmission can be used worldwide (cf Fig. 6.31).

Multiple access A promising application of RFID systems is the simultaneous recording of many transponders and thus of tagged objects. This multiple access also known as *bulk scan* is already supported by the radio technology since all transponders within the sphere of a scanner can send and receive data at the same time. While during a *broadcast* from the scanner all transponders simultaneously receive a data request, the response telegrams of the single transponders have to be recorded and processed collision-free by the scanner by means of adequate procedures.

Using access methods known from radio technology, one can distribute the capacity of one transmission channel to several participants (multiplex operation). However, this can only be used in case of a continuous radio transmission. Passive read-only RFID transponders with little data volumes transmit for short, discontinued periods of time and with long breaks. In

[29] Metals or liquids in the sphere reduce the strength of the radio signals emitted by an RFID system.
[30] Cf. CEPT/ERC REC 70-03 174, for Germany EN 300 220-1, EN 300 220-2, EN 300 330 and EN 300 440.



Source: ETSI

Figure 6.31. Frequency ranges for RFID systems (worldwide)

such cases, multiplex methods cannot be used or only to a limited extent. *Anti-collision methods* are used instead which optionally are *transponder* or *scanner-controlled*:

ALOHA: This simplest transponder-controlled method makes use of the differences between the repeat times of transponders. Since the transmission time of a telegram is much shorter than the repeat time until the next telegram and since all transponders with a defined possibility send at different times the collision rate and thus the average time required for the reading of all transponders can be estimated[31].

Slotted ALOHA (S-ALOHA): Is an optimization of the above-mentioned method by a synchronized data transmission where the transmitter determines time slots for all transponders by means of a broadcast telegram.

Dynamical S-ALOHA: In this method the number of slots initiated per time unit can be varied to increase the reliability of the scan.

There are further methods as well as error recognition techniques. In addition to this, safety mechanisms have to be considered in certain fields of application which can be used to encode the transmitted data [1, 10].

[31] Finkenzeller sets 800 ms as an average time for the reading of eight passive transponders. To ensure a reading reliability of 99.9% about 2.7 s would be required.

6.7.2 Fields of application

Tagged transport packaging opens up all fields of application for barcode systems. Furthermore, the direct automatic scanning and booking, for example, during a bulk scan, allow for the consistent tracking and tracing along the complete supply chain without manual interference.

General use in logistics The tagging of goods and packaging in the consumer goods industry with transponders offers a large potential for the optimization of the supply chain. But because of lacking standards and relatively high costs this method is only used in special cases or for high-quality goods.[32]

Principally, transponders can be used in all areas of automatic identification, control and partially for the recording of measuring data[33] This method is ideal for reusable packaging, such as containers, crates, barrels, etc. Since transponders manage the product-specific information as well as information about the owner, times of use and cycles.

They can further be used to protect the originality of products (forge-proof) and to ensure the consistent traceability along the production and supply chain.

RFID in commerce — ePC The *Electronic Product Code* (ePC) was developed in 1999 by a cooperation of the Massachusetts Institute of Technology (MIT) and different American trading companies with the aim to standardize and propagate the ePC. The resulting code standard provides a data space for transponders of 96 bits by which about $7,9 \cdot 10^{28}$ unique article codes can be assigned. Transponder costs should be minimized by means of the small memory. Instead of storing the article data and data histories locally the code serves as a key by which the article-related data can be retrieved from the Internet.

The ePC is implemented on transponders physically in the microwave or UHF range. On request each code is assigned to a storage location for the article data by means of *object name services*[34] (ONS). A standardized semantics is ensured by the *physical markup language* (PML).

Since all information is highly decentralized, all data flows in the Internet are controlled by so-called *savants*[35]. These distributed software systems collect, count and filter all transponder numbers arriving in the network, organize the data update between the distributed systems and offer interfaces for the local connection of a scanner [27].

---

[32] For example for computer components (hard disks, memories) or media (sound carriers).

[33] For example, Smart-Labels with an integrated temperature sensor to control food along the cold chain.

[34] Cf. *Domain Name Service* (DNS) an Internet service which translates requested URLs into IP addresses.

[35] cf. `www.autoidlab.com`

6.7.3 Comparison with barcode systems

The characteristics and applications of barcodes and RFID systems described earlier should now be used to evaluate the two techniques. Fig. 6.32 compares both systems by means of their application characteristics without consideration of specific differences like the rewritability of transponders.

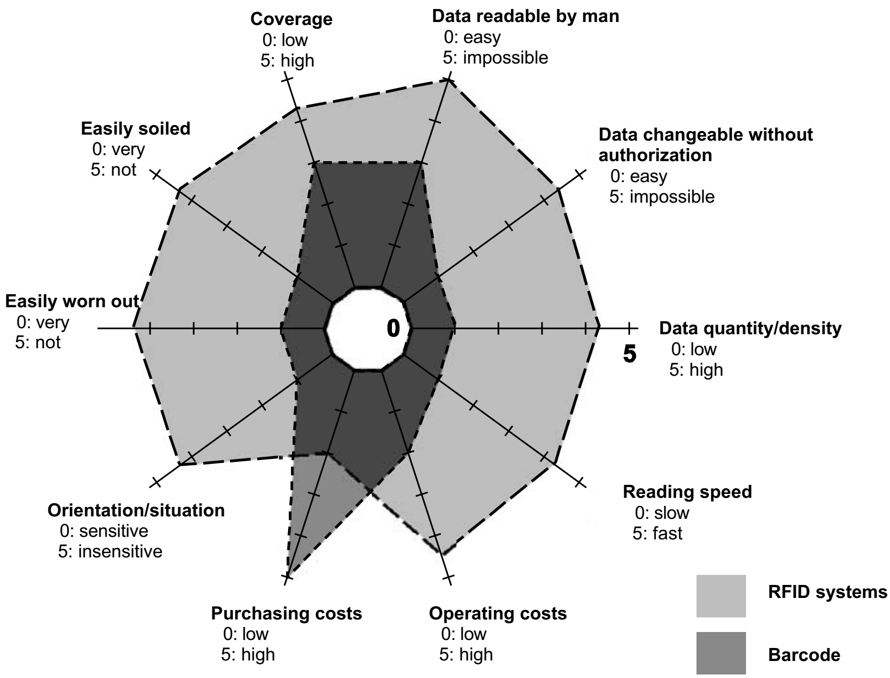


Figure 6.32. Comparison of RFID and barcode systems

7. Information and Communication Technology

This chapter deals with the basics of information and communication technology with the aim to give a better understanding of general principles and their practical implementation – as far as they are relevant for decisions about the choice, procurement and operation of a WMS. Since the information and communication technology is subject to rapid changes this chapter just gives a basic overview and describes concrete procedures by means of chosen examples.

7.1 Communication technology

Modern communication technology is the key to a quick and faultless data exchange and to various services offered in public networks [76] [77]. The era of electronic communication started when the first Morse telegraphs were sent. These were soon followed by telephone, telefax and television as well as modern digital technologies. The transmission techniques have also changed a lot. While at the beginning only wired point-to-point connections without relays were possible these were followed by switched connections[1], packet switching[2] and broadband technology[3]. Furthermore, the multiple use of media by means of the multiplex technology, the optical transmission and satellite technology opened up new possibilities. Today, a variety of services are offered for different transmission media.

The following sections give a short overview over the many terms and abbreviations related to communication technology.

[1] The switching technology where a line is switched for the time of a connection is called *line connection*.

[2] The data to be transmitted are combined into small units, packets. Each of these packets is transported individually through the network instead of sending the complete data over a fixed line.

[3] The same frequency band in a transmission channel is used simultaneously for several data transmissions.

7.1.1 Layered architectures

Layered architectures represent just *one* software structure. In communication technology this principle is used with great success. A *layer* provides services for the superordinate layer using the services offered by its subordinate layer (cf. Fig. 7.1).

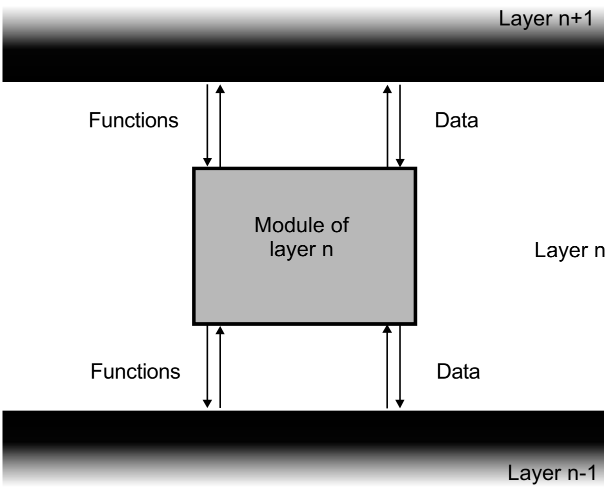


Figure 7.1. Basic principle of a layer in a software system

This and the following sections describe the *ISO/OSI reference model* (cf. Fig. 7.2) which is widely used and recognized in the communication technology. With the ISO/OSI model the communication via electronic media can be represented in seven hierarchical layers, which represent the information flow from the physical medium to the application (at the example of a WMS). Applications are connected with each other via the layer sequence $6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow$ transmitter $\rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$. Since on the receiving side the layers are passed in reversed order it is also called a *protocol stack*. Table 7.1 shows the tasks of the single layers. The layers 1 to 4 are considered as transport-oriented layers while the layers 5 to 7 are application-oriented. A detailed description can be found, for example, in [77].

7.1.2 Protocols

Protocols implement the functions on a layer. Each single layer receives the active data to be transmitted from its superordinate layer. Together with

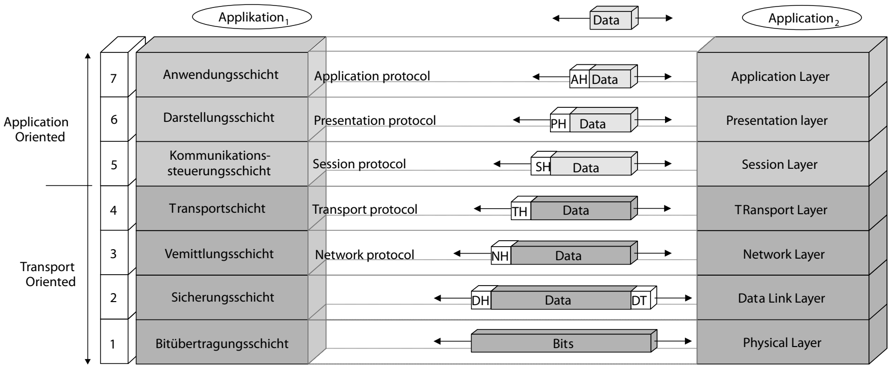


Figure 7.2. The ISO/OSI reference model. The English terms are shown on the right of the protocol stack, the German terms on the left

the *header*, which contains information about the transmitted data, these data are combined into a *frame* or *packet*[4] and transferred to the next layer. During this procedure larger packets can be split up into several smaller ones or combined into larger packets. On the receiving side, the headers generated on the different layers are separated from the active data so that the application only gets the data sent by the sending application.

Received packets are in most cases accepted or rejected by acknowledgment packets which, by means of checksum calculations and evaluation protocols, identify transmission errors and can request the corresponding layer to resend the packet. For a detailed description of general protocol principles please refer, for example, to [35, 77]. Typical protocols used in the Internet are shown in Fig. 7.3:

- Internet protocols
  - *FTP* (file transfer protocol) for the secure transmission of files
  - *TELNET* (remote terminal program) for a character-based communication with system and application programs
  - *SMTP* (simple mail transfer protocol) for the transmission of emails
  - *NSP* (name server protocol) for the breakup of names in the Internet
  - *SNMP* (simple network management protocol) for network management
- Transport protocols
  - *TCP* (transmission control protocol) for the connection-oriented communication. A client process (cf. section 7.1.6) connects to a server process. This connection can be used for communication in both directions until it is terminated by one of the users.

[4] Frames and packets always include a header and active data. The term "frame" is mainly used for the layers one and two.

Table 7.1. Layers of the ISO/OSI reference model and their tasks

No.	Layer	Task
7	Application	Application program which communicates with one or several other applications using the protocol stack.
6	Presentation	Standardization of a representation of data types in the network independent of the application. For example, integer numbers are represented in the network always in a certain order while in an application the representation depends of the computer hardware.
5	Session	Here, status data are stored for the runtime of an application. So-called synchronization points are managed above all for troubleshooting purposes.
4	Transport	The transmitted data are split into packets. The order of received packets is ensured.
3	Networking	The packets which are not destined for this computer are transferred to another computer. This networking may include several computers and is called *routing*. The computers which use this layer only as node between neighbouring computers and do not transfer received packets to a higher protocol layer are called routers (cf. Fig. 7.11).
2	Data link	Securing of the correct data transmission. The correctness is also checked on the higher layers but this layer is the first one with a direct connection to the hardware where a high error rate has to be expected depending on the used transmission medium and the access procedure as well as the current error situation. The smallest unit of transmitted data is called frame. This layer is divided into two sub-layers, the medium access control (*MAC* layer: MAC regulates the access to the medium. It deals, for example, with collisions in bus systems (cf. Chapter 5). The LLC layer (*LLC*: Logical Link Control) is responsible for the secure data transmission. Check sums are added to the frames; acknowledgement frames acknowledge the correct reception of frames.
1	Bit transmission	Specification of physical requirements such as cables, plugs and voltage levels.

  – *UDP* (user datagram protocol) works connection-free. Each data packet is sent to the receiver independent of the transmitter. UDP gives no guarentee that the packets actually reach their destination.
– *IP* (Internet Protocol) as an example for a connection protocol. Often used in connection with the TCP transport protocol (→*TCP/IP*).
– Data link protocol according to IEEE802.x
  – *IEEE802.3* Ethernet as an example for a bus structure.
  – *IEEE802.4* Token bus as logical ring structure on a physical bus.
  – *IEEE802.5* Token ring as an example for a ring-based network structure.

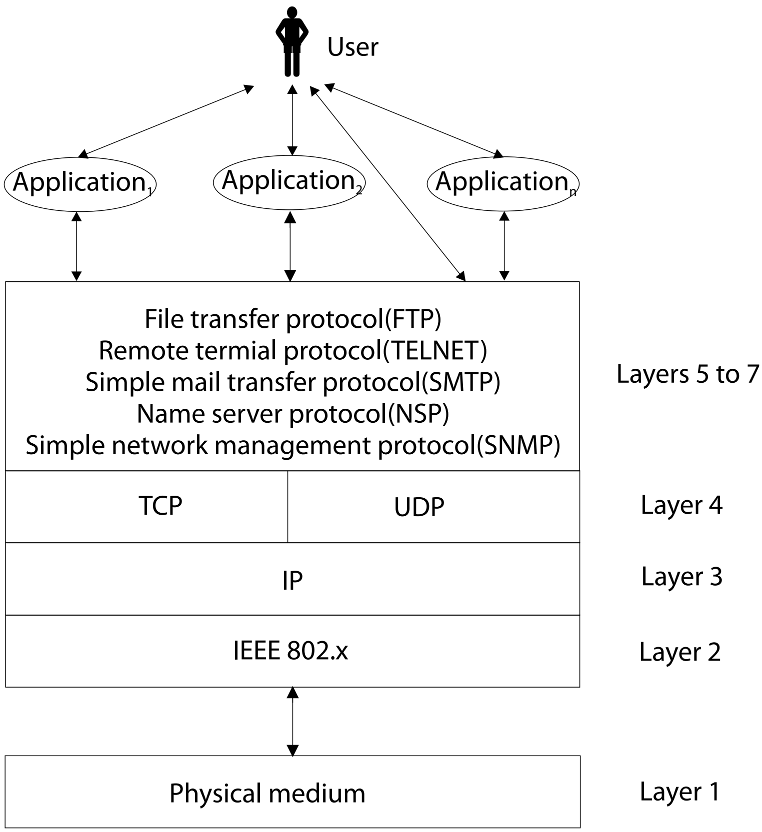


Figure 7.3. Typical protocols used in the Internet [77]

- *IEEE802.6* Optical double bus systems to be used in MANs.
- *IEEE802.11* Application is wireless LANs.

7.1.3 Transmission media

The transmission media are the bottom layer of a communication system. Despite of their analogous physical functioning they are used for the transmission of digital signals – i.e., signals with a finite countable set of symbols. On the higher layers only digital signals with two status – binary signals – are used. The electric cable, for example, can be charged at the same time with different voltages. The higher the possible voltage the more critical the assignment of a concrete variable to a symbol under real conditions. Under real conditions the voltages are always subject to fluctuations caused by the characteristics of the transmission medium and external interferences. For this reason the quantity of information to be sent at a certain time depends

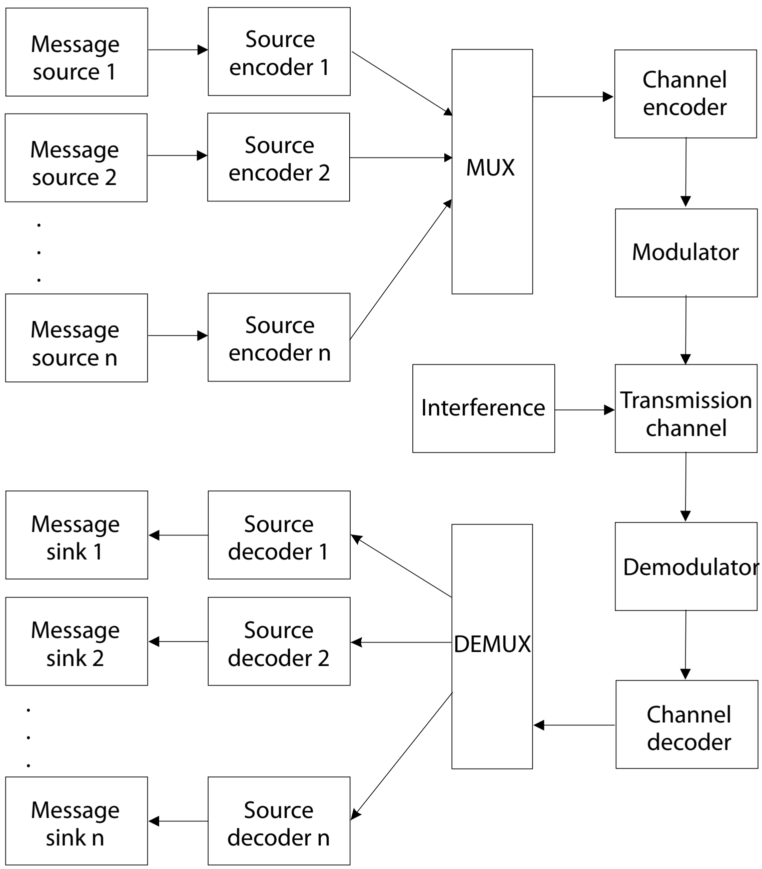


Figure 7.4. Source code and multiple access on the transmission medium

on the interferences and the quality of the medium. Many modems[5] which are used for a data transmission via telephone lines adjust automatically to the current transmission quality.

The transmission of voltages is just *one* way to transmit information to a medium. Alternatively, they can be transmitted by sounds, sound modulations but also by combined methods[6].

With optical fibres (also called glass fibres) signals can be transmitted nearly error-free. They cover a large band width and are insensitive to electromagnetic fields. Optical fibres are more and more used in all industrial networks. Figure 7.4 shows the basic structure of a data transmission from the source to the sink. The source code packs the data flow. The methods de-

---

[5] A modem consists of the components Modulator and Demodulator.

[6] Corresponding communication techniques (amplitude, frequency, phase and quadrature modulation) are described in [42].

pend on the kind of data source. Digital images and audio files, for example, are encoded by other methods that text files. To better utilize the transmission channel several data flows may use the same transmission medium. This technique is called multiplexing, the corresponding devices are multiplexers/demultiplexers (MUX/DEMUX). The channel encoder optimizes the bundled signals according to the characteristics of the transmission channel [42]. Adaptive methods work dynamically and consider the current interferences. When the interferences acting on the transmission channel increase the transmission speed is reduced. The modulator converts the signals into physical values (electrical/optical) which are then sent through the channel to the destination. This basic diagram does not show the bi-directional use of the channel but just point-to-point connections: Each data source is assigned to just one data sink. Section 7.1.4 shows other possible network structures and connections.

Another important factor is the transmission speed. The number of figure-shift signals per second is described by the unit *baud* and is restricted by the characteristics of the medium (*bandwidth*). For the user the *transmission rate* in bits per second is more important since this value has a direct effect on the transmission time required for a given data volume.

When longer cables are used the signal lapse is *slurred* and the signal strength is reduced. For this reason, repeaters (cf. section 7.1.4) are used to restore and increase the signals.

In addition to the electrical transmission media described in this chapter also optical media are used which function independently of external electromagnetic fields and emit no such fields themselves, i.e., have a high electro-magnetic compatibility (*EMC*). They can be laid together with other cables without causing any problems.

Special attention has to be given to radio transmission. Above all shading, large signal fluctuations at the destination and multiple input due to reflections complicate the technical implementation of this technology. Especially in warehouses with a lot of steel elements, the use of radio transmission has to be planned and set up carefully and often requires extensive measurements and, if necessary, additional transmitters and antennas.

7.1.4 Network types and internetworking

Networks can be categorized into bus, loop, double loop, hub-and-spoke, grid or complete systems according to their structure (cf. Fig. 7.5). With the exception of the bus structure the topologies are based on the point-to-point connection. In bus systems the communication between the users is somewhat difficult because data can be transmitted by just one user at a time. In Chap. 5 some methods (medium access methods) are described by which this problem can be solved. Similar problems occur in radio and infrared systems where several users share one medium. Because of the mobility of the users

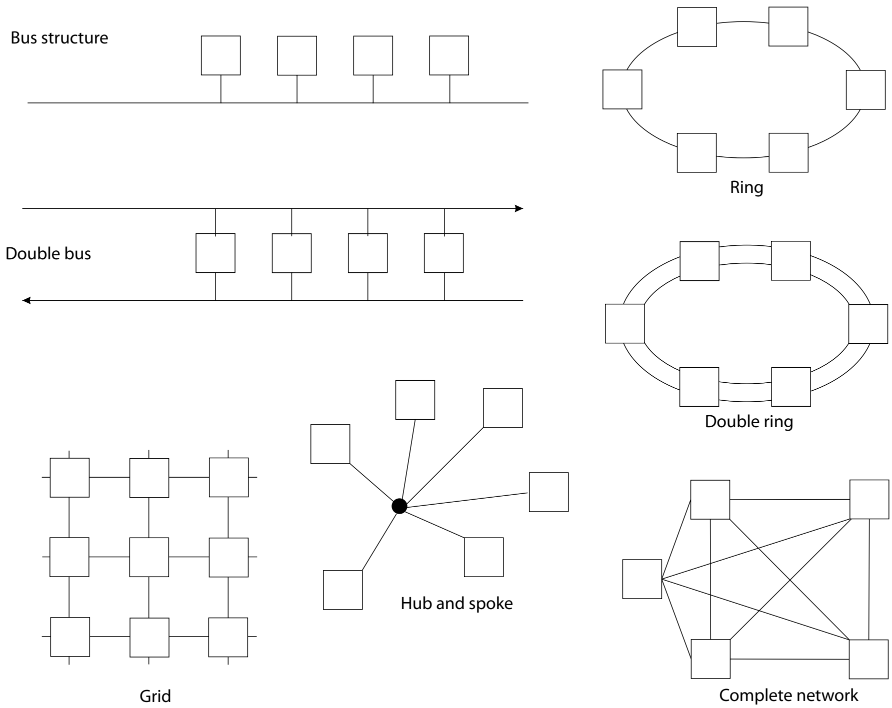


Figure 7.5. Typical network topologies

and the restricted ranges localization and highly dynamical switchings are required which are not described here [77].

From the field bus to long-distance traffic Networks are used in all industrial sectors. They are classified according to the length of the transmission line between two communication partners[7]. The error rate and latency increase in line with the distance. The networks are categorized as follows according to the typical distance between the single front-end computers:

– In automation equipment *field bus systems* (cf. Chapter 5) support the quick exchange of signals and measured values between transmitters and controls as well as between controls and actuators. The call for a short *latency* may range from a very short to a guaranteed maximum latency. The realtime behavior of the overall system has to be ensured also in the relatively slow conveying processes.
– *Local Area Networks* (*LANs*) are mainly used within buildings and achieve a high throughput at a short latency.

[7] This also applies to the distance between to front-end computers by which the data packets are transmitted to the target computer.

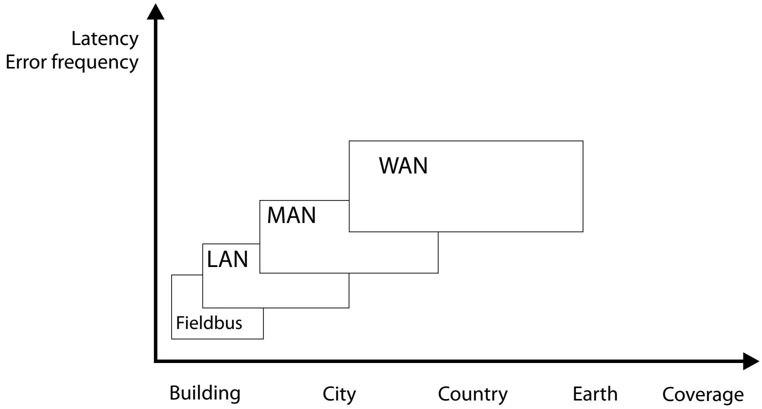


Figure 7.6. Rough classification of different networks

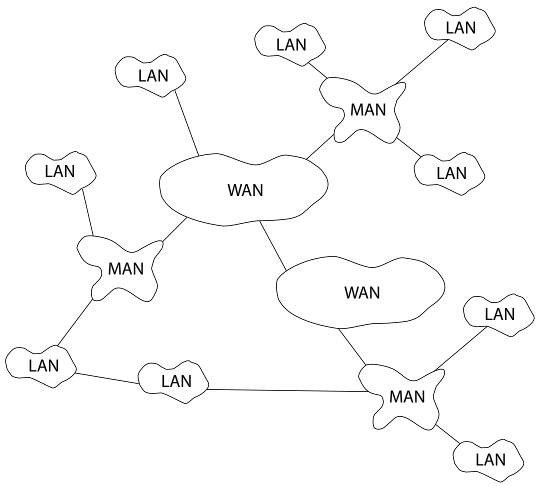


Figure 7.7. Combination of LAN, MAN and WAN into an internetwork

– *Metropolitan Area Networks* (*MANs*) are typically used for "medium" distances
– *wide area networks* (*WANs*) are used to bridge large distances – e.g., between cities, countries or continents.

Fig. 7.6 shows the qualitative classification of a network. Based on these networks and their characteristic features large networks – so-called Internet networks [77] – can be built. A typical example of this is the Internet. Its main connection lines, where large data volumes are transmitted over long distances, are called *backbones* (cf. Fig. 7.7).

Internetworking In a heterogenous world of communication large distances

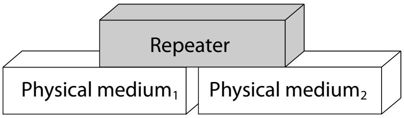


Figure 7.8. Repeater regenerating signals between cable segments

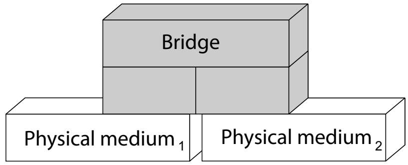


Figure 7.9. Bridges for the connection of networks with different backup protocols

have to be bridged and different networks have to be connected. The routing in such networks requires adequate, cross-network switching facilities. Subnetworks can be connected on the different levels of a layered system.

- *Repeaters* belong to the ISO/OSI layer 1 and regenerate electrical signals. They are used to bridge larger distances.
- *Hubs* also belong to layer 1 and are mainly used to reduce the cabling. They connect several segments each of which is a bus system of its own. A hub copies all incoming frames of each segment into all other segments. Thus the network has the topology of a hub-and-spoke system but behaves like a bus system (cf. Chapter 5).
- *Switches* are used to avoid the collision of the frames of different segments. Switches are structured like a hub but copy the received packet just into the segment connected to the receiver (also called *switching hubs*).
- Many companies operate several LANs which have to be connected. For this purpose *bridges* are used. Compared to a router which operates on layer 2 it requires less expenditure. Bridges can connect networks with different technologies – or better with different medium access protocols. The target address of the switching layer is not decoded for this purpose.
- *Routers* transmit the data packets in layer 3 via “favourable” paths.
- A *gateway* operates on the application layer and has to pass all subordinate layers of the protocol stack. Gateways are used to connect different applications or different subordinate protocol layers where a transition between different networks is possible only on this level. Gateways are often used to couple WMS and ERP systems to allow for a communication between these systems despite of their different application protocols.

7.1.5 Network addresses

Each layer of the ISO/OSI protocol stack operates with clear identifications which are called function code, session ID, service ID, node address or hardware address, depending on the layer. The addressing is described by a simple example[8]:

[8] The example is based on a TCP/IP protocol and an Ethernet bus system.

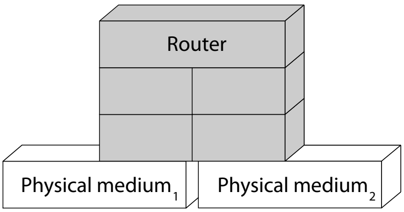


Figure 7.10. Diagram of a router for the transmission of data packets to the target hub

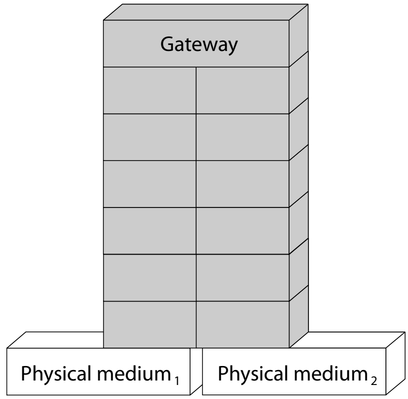


Figure 7.11. Gateway for the coupling of different applications

On layer 1 the physical connections are definitely identifiable and labeled units. Ethernet connections are typically labeled with eth=, eth1, etc. while RS232 connections are often labeled with com1 and com2[9]

The addresses in layer 2 are called *MAC addresses*. They are indispensable in bus systems for a clear user identification. The Ethernet operates with a worldwide defined 48-bit address which is assigned to a network card by the provider.

The addresses in layer 3 – the network layer – are similar to telephone numbers. Each user needs a clear ID independent of the technology and the protocols of layers 1 and 2. Figure 7.12 shows the often used IP address scheme ipv4 with a 32-bit address[10]. To improve the readability the addresses are represented by four numbers separated by dots — each in the range of 0 to 255 — and are assigned by the *IANA* (Internet Assigned Numbers Authority). The addresses

10.0.0.0 ... 10.255.255.255
172.16.0.0 ... 172.31.255.255
192.168.0.0 ... 192.168.255.255

play a special role since they cannot be managed centrally. They can be used in networks without Internet access. Addresses beginning with the bit pattern "1110" are used for so-called multicast addresses which are reserved for group communication (one sender – many recipients) [77]. The classification is very

9 Under Unix operating systems and under Linux the serial interfaces are labeled with 66yS0 and ttyS1.

10 Here, only the ipv4 will be described which with 2^{16} possible addresses is much too small for a worldwide use. For this reason procedures have been developed for the dynamical assignment of IP addresses. Currently, the ipv4 is changed to ipv6 with 2^{128} possible addresses.

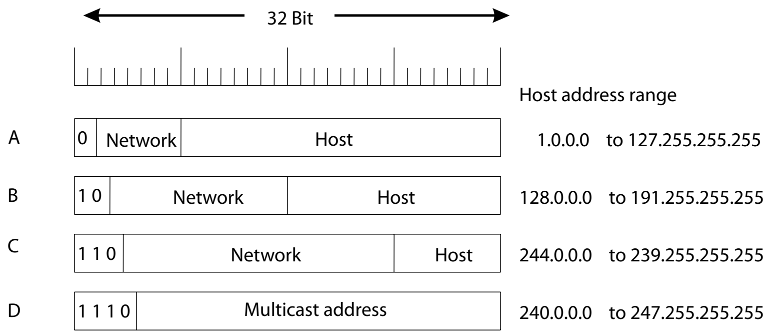


Figure 7.12. Simplified addressing scheme of the network protocol IP in the version ipv4

rigid and networks of the classes A 2^{24}, B 2^{16}, and C 2^8 can include users. By means of a so-called *network mask* by which partial areas of the address can be marked, this address scheme can be adjusted to better meet the daily requirements.

Layer 4 uses an ID defined for each user or computer node. Under *TCP* (transport control protocol) this is called a *port*.

Layer 5 uses a *session ID* for the time of the connection by which the user can be identified. Layer 6 identifies the data types via corresponding IDs and layer 7 can use function codes or object IDs as defined ID depending on the used application.

7.1.6 Examples

Client-server model Networks can be used to provide services to other users. Such a provider is called *server*, the user is called *client*[11]. A file server, for example, gives access to a file system. These services can be used by several clients *concurrently*. Such an architecture has to be able to deal with different kinds of errors. For example, the server, one of the clients or the network may break down in different system stati [77]. A process may be server and client at the same time when it uses the services of another server to provide its services.

WMSs often use the client-server principle to manage, for example, decentralized workstations. The inventory data are stored centrally on a server while warehouse-specific applications run on the client using local capacities. Examples of such a system are identification stations at the I-point, order

[11] The client-server principle is universal and bound to a network. A server can also be implemented on a single computer in the form of processes while other processes represent the client.

management for the entry and control of retrievals and mobile terminals in the order-picking area which work with stationary or local client programs.

Time synchronization at the example of NTP In a WMS which consists of several computers and/or is coupled online to other systems the consistency of the system can only be ensured by use of a time protocol (cf. section 7.4.2). The *NTP* (network time protocol) is *one* possibility to synchronize the clocks in different networks. Basically, the average value is built and a distributed algorithm is used which runs on different computers.

Name service DNS The *DNS* (domain name service) is an Internet service designed for assigning hierarchically structured identifiers to a number of characteristics [77].

The hierarchy is specified by dots which are separated by a sequence of identifiers where the rightmost represents the highest level – the so-called top level domain. The name can be refined by adding one dot and one identifier at the left. The part of the identifier to the right of the dot is called "domain", the part left of the separator is called a "name", which has to be unique only within its domain.

The most important application is the assignment of an IP address to an identifier. Thus it is possible to specify a computer in the Internet by a "talking name" instead of the numerical IP address which often is difficult to memorize. Furthermore, the numerical address may change during a reorganization while the name remains the same. Aliases referring to the same computer can also be used[12]. The DNS can further be used, for example, to assign computers to mail servers or to display the computer in plain writing.

In the Internet DNS is a distributed data base. The characteristics of an identifier – for example the assigned IP-addresses – are decoded by a *name server*, which is permanently informed about the status of all computers in its domain. To decode external addresses the name server uses the services of other name servers and stores these information for a certain period of time to reduce the data traffic.

URL The requirement for a standardized addressing scheme is met by the *URL* (*universal resource locator*)[13]. The hierarchical name space of a URL is built by combining several name spaces. The first is the protocol identifier which is terminated by a colon. The specified protocol, on the one hand, determines the used application protocol and, on the other hand, the further structure of the URL. This often begins with a slash when file names are specified and with a double slash for computers. Computers can be specified optionally by their name (cf. section 7.1.6) or their IP address (cf. section 7.1.5).

[12] www.mywms.de and www.mywms.com refer to the same IP address.
[13] The term *URI* (Universal resource identifier) is used synonymously but is not as widely used [99].

Table 7.2. Examples of URLs

URL	Description
http://mywms.de	*Hypertext transfer protocol* for the *standard port* of the HTTP server of the computer *mywms.de*.
http://mywms.de	*Hypertext transfer protocol* for the *port* 8080 of the http server of the computer *mywms.de*.
mailto://info@mywms.de	Send an *e-mail* to the computer *mywms.de* for the recipient *info*.
ftp://dante.org/tex/v1.tgz	*File transfer protocol* to read the file */tex/v1.tgz* sent by the computer *dante.org*.
ftp://alice:bob@dante.org/tex/v1.tgz	As above, but the user logs in with the name *alice* and the password *bob*.

WWW The WWW (World Wide Web) is an Internet service, which was developed in 1989 by CERN in Geneva with the aim to standardize the representation and connection of documents. Meanwhile, the WWW has become the most popular Internet provider and the Internet is colloquially called WWW or just the Web. The service functions according to the client-server principle where the server manages the documents which are connected via URLs or so-called *links*. A link may refer to parts of the documents in which it is defined or as external link to other documents stored on external computers. The documents are stored and transmitted in a *logic* format in the hypertext markup language (HTML). The exact type of representation is determined by the client. The representation tool for HTML documents, the *browser*, adapts the received documents to the user requirements by means of local settings. These requirements are, for example, font size, color or page dimensions but also security settings like the linking of received documents to applications. The documents are transmitted in form of a simple protocol in plain writing[14], the *HTTP* (*hypertext transfer protocol*).

Documents are not limited to plain text but also appear in a variety of other forms like multimedia documents or application-specific documents, e.g., drawings or article master data.

Documents are specified by their *mime types* (*multipurpose Internet mail extensions*) [77], which originally were developed for emails (cf. Table 7.3).

In WMS HTML is more and more used for user interfaces because owing to the standardized communication the terminals just need a suitable browser.

[14] *Plain writing* means that only printable characters are transmitted.

Table 7.3. Some examples of MIME types

Type	Sub-type	Description
Text	Plain	Unformatted text
	Richtext	Text with simple formatting
Image	gif	Image in graphics format "gif"
	jpg	Image in graphics format "jpeg"
Audio	wav	Audio file
Application	octet-stream	Application-specific byte-flow

These are already installed on many PCs in the offices and mobile terminals with HTML browsers are more and more available on the market.

Release of resources This network technology offers the possibility to provide local devices, directories and files on the net to be used by other users. Advantages of this multiple-client system are cost savings and the use of central services like the centralized backup of a file server.

7.2 Data management

The data management is an important aspect of the warehouse management. In addition to the article and inventory data – the basic data of a warehouse management system – quite a number of other data have to be managed. These data are differentiated according to their dynamics. Statical data, which can be changed during operation, are, for example, the number and dimensions of the storage bins[15]. Data with little dynamics are article master data because they are only generated when new articles are entered and changed when the article characteristics are changed. Highly dynamical data are, for example, order and inventory data because they change with every movement in the warehouse (cf. section 2.4).

The following sections focus on the safe and permanent storage, the coherence of all data in a system, the quick access and filing of data. They also deal with the relational database which nowadays is the most common standard for the data management in complex systems.

7.2.1 Principles

The following principles and definitions are independent of the type of data management.

[15] Some systems allow for such changes during operation but the securing of the system integrity may be very costly.

Persisting The permanent storage of data independent of the method is called *persisting*. Programs store their data during the runtime so that they will not get lost in case of short circuits, hardware defects or program errors. Sometimes it is necessary to cancel a running system for maintenance purposes and to restart it later on. In such cases the data have to be stored permanently, i.e., they have to be persisted.

Referential integrity Data must not be deleted when they are still used by other parts of the program (referenced). This *referential integrity* ensures, for example, that article master data in a WMS are deleted when there still are corresponding articles in the warehouse or announced by the supplier.

Transactions During the data processing temporarily inconsistent states occur from time to time. When, for example, a new pallet enters the system the data have to be stored, the pallet counter has to be incremented and a location bin has to be reserved. Since the program flow is sequential this leads to temporary inconsistent states. This problem can be solved by *transactions* which "clip" such operations. Each state before a transaction is consistent as is each state after a transaction. Transaction systems offer the possibility of a *rollback* to cancel a transaction and to return to the former state. A rollback deletes all changes performed during the transaction. The opposite – the successful conclusion of a transaction – is started by a *commit* so that the changed data are stored persistently and the system returns to a new consistent state.

Access and blocking mechanisms The concurrent use of data and the transaction concept require access mechanisms which dependent on the planned operations. A blocking may refer to just the read data or to complete files as well as to database tables or only to parts thereof. Some data management systems offer the option to read also blocked data. A user of this *dirty read* principle has to bear in mind that these data may be changed concurrently. This type of access is used, for example, for inquiries.

Trigger Changes of data sets often initiate further actions. The basic principle that a status change becomes an event is called trigger – in this concrete case a *data trigger*. In a WMS the change of an article feature leads to an update of other data. These facts can be encoded in the program functions but data triggers *always* ensure that these changes are made and prevents potential programming errors.

Journaling The persistent storage of large data sets is very time-consuming and in some cases it may be better to save the operations instead of the data after a data set has been changed. This principle is called *journaling* and is widely used – also for data storage. Transactions and "undo functions" are mostly performed by journaling. Starting from an initial status all changes are stored in a so-called *journal file*. After a system breakdown the current status[16] can be restored by means of the operations stored in the journal file. In journaling the persisting thus only refers to status *changes* and not to the status itself. To keep the journal file small the current system status has to be persisted from time to time and the journal files have to be deleted.

Filing In contrast to the persisting data *filing* is used to save data for a long term. It is important for the future evaluation of data and above all for the preparation of reports and other offline evaluations. There may even be regulations for the filing of data or they have to be filed for quality assurance, e.g., the filing of delivery notes and inventory data.

7.2.2 File systems

Each operating system offers the option of a persistent data storage[17] in the form of a *file system*. Files are referenced by names which generally are hierarchical. This hierarchy is represented by *Directories*. A file can be referenced by an absolute or a relative *path*. Each hierarchy begins with a root and ends at a file. The root of a file system can start with the memory of the file system or with a logical identifier(UNIX: "/", WINDOWS: "a:, b:, c:, d:, etc."). Special characters are often allowed but are interpreted differently by some systems. These aspects have to be considered especially in distributed and portable WMS. Here, special characters and blanks should be avoided and all file names should be written in lower cases. Recommendations have been established for portable file identifiers.

The logical structure of a file determines how certain data sets can be selected:

- *Sequence access*: The data sets of a file can be read and written only sequentially. This kind of file organization is used, for example, to write time series for statistical evaluations. Because of their physical structure tape machines support just this kind of access.
- *Index sequence access method* (ISAM): Each data set can be selected by means of a key. Then data can be accessed sequentially from this position. In the past ISAM files were often used in warehouse management systems. Today, this access method is used by databases and thus is hidden from other applications.

[16] Open transactions are not carried out during the restorage so that after a system breakdown, it has always to be checked if the last operations have been performed.

[17] So-called RAM disks in the central memory of a computer represent a volatile file system and is of use when an application requires the functions of a file system except the persisting. Because of the short access time, RAM-disks are ideal for tests. Generally, they are not used in WMS.

Table 7.4. Typical file attributes and their meaning

Attribute	Meaning
Read	File contents can be read
Write	File can be written
Delete	File can be deleted. In many systems this corresponds to the permission to write into the file directory.
Execute	The file can be executed as code. This may be a file with an executable machine code or, in some systems, an interpretable code.
Lock	The file is already open and cannot be opened anew
Archive	This file has to be filed in a backup
Password	Password required for access to this file
Owner	Owner of this file
Date of generation	Time and data of the generation
Date of last access	Time and date of the last reading access
Date of the last change	Time and date of the last writing access

– *Random access*: Each data set can be selected by means of an integer number. Thus, an application can access any files by a so-called *key transformation*. In the past, data sets related to the bin locations were selected by a random access. Present systems mainly use a database for this purpose.

The access on files is controlled by a *safety mechanism* which in most cases consists of a correction system for the owner, the group members and other users.

By *encoding* the data content it can be further protected against misuse – should they be meddled with.

7.2.3 Databases

Today the term *database* generally describes a relational database (*RDBS*: Relational Database System) the basic principle of which is described in the following section. The tasks of a database are:

– to store data persistently
– to ensure the integrity of the data
– to give access to data sets
– to link data sets
– to save, change and delete data sets

In addition to this they meet the general data management requirements described in section 7.2.1. Chapter 9 gives an example of a WMS based on a relational database.

The real or imaginary elementary data sets managed in a database (records, tuple) are called *entities*. The structure of the data set is determined by attributes with the corresponding attribute values. Each attribute of a data set is assigned to a data type. Similar structured data sets are stored in a *table* The data sets in a table have to differ by at least one attribute value. Thus, a table, in the mathematical sense, is a *relation* (s. Table 7.5).

Table 7.5. Representation of a relation in the form of a table

Relation ARTICLE   Attribute ↓

Number	Name	Weight
3973684	PU-foam	0.800
3974954	Bitumen	30.000
3978617	Spackle	5.000

← Tuple

↑ Attribute value

The relations between entities are described by direct *associations*. An association $EM_1 \leftarrow EM_2$ determines how many entities of EM_2 *may* be assigned to an entity of EM_1. In practice, only four basic kinds of associations are differentiated. Table 7.8 gives examples of such associations.

Table 7.6. Association types and their meaning

Association type	Number of entities of EM_2 assigned to *one* entity of. EM_1
1 : simple association	1
c : conditional association	$c \in \{0, 1\}$
m : multiple association	$m \geq 1$
mc : multiple conditional association	$mc \geq 0$

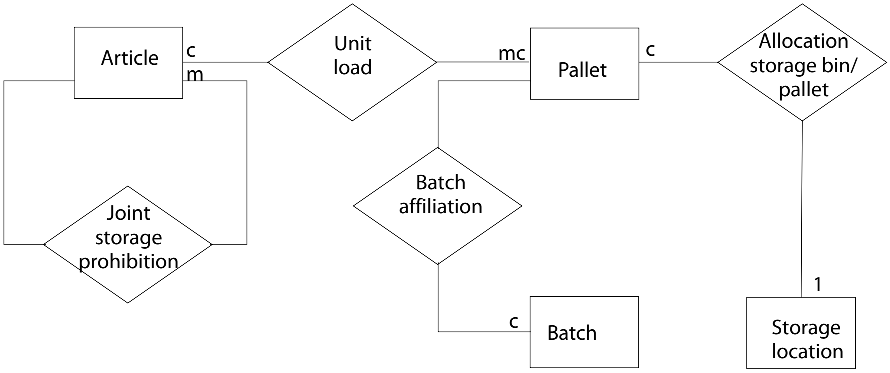


Figure 7.13. A typical entity-relationship diagram at the example of some aspects of a warehouse management

Bi-directional associations can be built from the elementary association types. Table 7.7 describes the hierarchical, conditional and the network associations.

Table 7.7. Classification of association types according to [106]

	1	c	m	mc	
1	1 – 1	c – 1	m – 1	mc – 1	← hierarchical
c	1 – c	c – c	m – c	mc – c	← conditional
m	1 – m	c – m	m – m	mc – m	← network
mc	1 – mc	c – mc	m – mc	mc – mc	

Each table needs a defined *primary key* which is assigned to just one attribute. Many databases also support segmented keys spanning several attributes. In order to speed up the access on single data sets *indexes* can be entered which may also be assigned to an attribute or a group of attributes. However, each additional index requires additional memory space and computing time.

The association between several relationships are represented by *entity relationship diagrams*. The association are relationships themselves which are implemented in databases also by tables. Figure 7.13 shows an example. The entities *article*, *pallet*, *bin location* and *batch* are linked by the associations *joint-storage prohibition*, *unit load*, *batch* and *pallet bin assignment*. In the diagram the association type is shown on the connecting line.

Up to now this book described a database from the logical point of view (logical schema). A database is defined by three *database schemes*:

Table 7.8. Examples of associations

Entity volume_1	Entity volume_2	Association typ	Name
Article	Article	m — m	Joint-storage prohibition
Article	Pallet	mc — mc	Unit load (mixed pallet)
Article	Pallet	c — mc	Unit load (single-article pallet)
Pallet	Bin location	c — 1	Pallet location (exact bin)
Pallet	Lagerort	mc — 1	Pallet location (exact section)
Charge	Pallet	c — m	Batch of pallet (condition: single-batch pallets)
Zone	Storage bin	m — m	Zone defintion
Bin location	Bin location	m — m	Distance (transport costs)

– The *logical* schema describes the table structure (see above).
– One or several *external* schemes describe *views* on the logical schema. Each schema contains a section of the views, which is needed by or allowed for the application, a user or a group of users.
– The *internal* schema describes the physical organization of the data storage and the access organization.

Before data are entered, read and changed in a database this has to be set up by means of a data definition language. This description defines the tables, their structure and types of attributes as well as the indexes. *DML* (data manipulation language) are used to operate in a database as well as for data requests and changes. In most database systems DDL and DML are combined in one language, the *SQL* (*structured query language*). With SQL tables, indexes and views can be built and data can be searched, linked, changed, inserted and deleted. Table 7.5 can be built with the commands

```
create table ARTICLE (
               NUMBER  number(7)
               NAME char(40)
               WEIGHT number(8) )
```

The values are entered into the table with the commands

```
insert into ARTICLE values (3973684, PU foam, 0.800)
insert into ARTICLE values (3974954, Bitumen, 30.000)
insert into ARTICLE values (3978617, Spackle, 5.000)
```

The basic operations which can be performed in a table are described by simple SQL examples. The result of such a link is always shown in a new table.

– With the *projection* certain attributes (columns) can be chosen from the table.

```
select NUMBER NAME from ARTICLE
```

The result appears in Table 7.5

```
3973684 PU foam
3974954 Bitumen
3978617 Spackle
```

The basic operations of the selection and the projection can be linked by an additional *where* clause.

– With the *selection* certain tuples (lines) can be selected from a table[18].

```
select * from ARTICLE where WEIGHT > 20.000
```

The result is shown in Table 7.5

```
3974954  Bitumen  30.000
```

.

– The *join* links the joint attributes of the two tables P and Q[19]. The result is a table with a new structure which contains the joint attributes of both original tables just once. The new tuple includes all combinations of the tuples of P and Q with identical attribute values. A join of the tables ARTICLE and UNIT LOAD is described in SQL by

```
select * from ARTICLE, UNIT LOAD
```

The result are all tuples or ARTICLE combined with all tuples of UNIT LOAD[20] with the same article number.

In addition to this *natural join* (⊗-join) there is the *Θ join*, which contains further regulations for the attribute values [5]. A Θ join of the tables ARTICLE and UNIT LOAD can be described in SQL by an additional where clause.

```
select * from ARTICLE, UNIT LOAD
              where ARTICLE.WEIGHT > 20.000
```

Other elementary links can be achieved by the operations (`union`), (`intersection`) and (`minus`).

Databases are client-server applications where the actual database, the *database engine* (*DBMS*: database management system) acts as server for the clients. These clients are applications which perform the data operations. In a WMS this may be, for example, a client for the maintenance of master

[18] The * symbol after the select means that the stated tupels contain all attributes.

[19] Two attributes are alike when their identifier as well as their values are the same.

[20] The table UNIT LOAD links article numbers with pallet numbers but is not shown here for lack of space.

data or for the goods receipt. The database engine is either accessed via library modules which are part of the database or via a standard protocol such as the *ODBC* (*open database connectivity*). With ODBC an application can link – possibly via a network – with a database to use its services. For this purpose a suitable ODBC driver has to be installed on the client which is generally included in the delivery.

In a client-server architecture of a database system each manipulated data set has to be sent by the server – the database – to the client where it is manipulated and then returned. Frequent or extensive operations of such kind – like the blocking of all shelves in an aisle – reduce the performance considerably. Therefore most databases offer so-called *stored procedures*. These are functions stored on the server which, after being activated by a client, independently carry out an operation without a data transfer between server and client. Owing to the quick performance the data sets are blocked only for a short time. This often results in a drastic increase of the overall throughput.

In addition to the relational database more and more object-oriented database systems (*ODBS*) come on the market. They allow for the direct positioning of software objects so that databases can also be used for object-oriented techniques.

7.2.4 Availability of data

The availability of data stocks must not be mixed up with a safe integrity or an access protection. These aspects are dealt with in the section Safety of data stocks. The current data have to be available at all times (online availability) and older data on request (offline availability).

Uninterruptible power supply The availability of data has to be ensured also in case of power failures and hardware defects. A power failure can be bridged for a short time by means of an uninterruptible power supply (*UPS*). During this time the power is provided by accumulator batteries. The computer is informed about this load alternation by – mostly serial – interfaces. If the line voltage is not restored after a certain time the operating system has to close all applications and to "shut down" the computer to save the data stock. The application programs have to shut down in a controlled manner when they receive such a signal from the operating system. Running transactions (cf. section 7.2.1) and open files have to be closed as well as the connection to databases.

Redundant data management Data protection in case of a hardware breakdown can be achieved by a redundant data management. These methods are known as RAID systems (redundant array of independent disks). The levels linear and 0 have no redundancy but where nevertheless included in the RAID scheme by the ANSI (American National Standards Institute) which has defined this standard. The most common methods are the levels 1 and 5. Levels 0 and 1 can also be used together. This method then is called level 10 or level 0/1. The RAID levels described here are as follows:

RAID 0 is also called *disk striping* and has no redundancy. It is only used to speed up the access and should therefore only be used together with one of the other levels. The data flow is divided into blocks and written parallel on the installed drives. If there are n drives the blocks 1 to n are written on the drives 1 to n, then the blocks $n+1$ to $2n$ also on the drives 1 to n, and so on in numerically ascending order. The data are read analogously. This method does offer no fault protection. If a driver fails all data on the other drives will be lost.

RAID 1 is also called *mirroring* and is based on the double storage of data (data mirroring). For this reason, only half of the disk space can be used. When a drive fails the operations are carried out with the remaining one.

RAID 5 adds a checksum to the data and distributes the data as well as the checksum to different drives. RAID 5 is ideal for frequent accesses with small data volumes. For this reason it is often used in transaction-oriented environments like databases. When a drive fails the data can be reconstructed on the remaining drives.

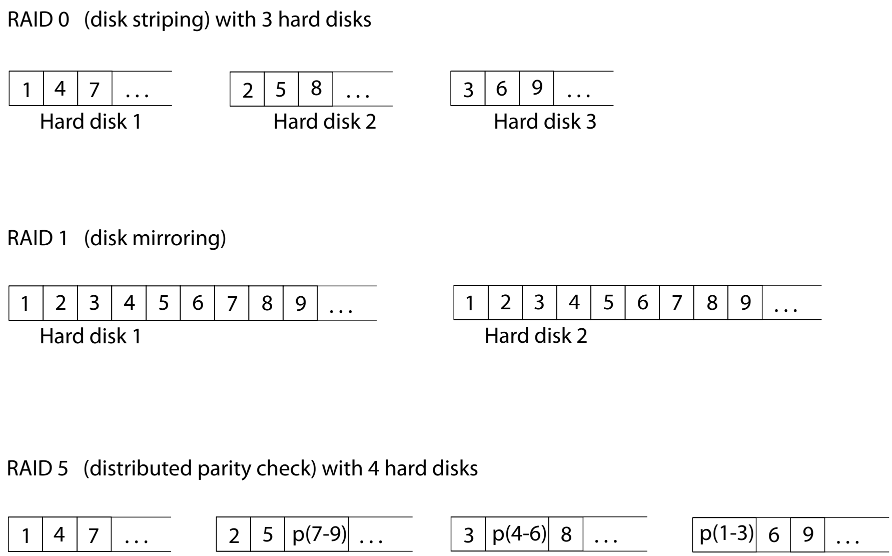


Figure 7.14. Simple diagram showing the storage of nine data blocks using different RAID levels

Journaling Another method to reconstruct data is *journaling*. All data changes are stored in a journal file which is written always before the data are changed. Based on the last backup the current status can be reconstructed by a so-called *recovery*. For this purpose, the journal file has to be available. The actual data file and the journal file are stored on different hard disks so that the loss of one of the two files does not cause too much trouble.

Backup *Backups* are used to reduce the damages caused by hardware breakdowns or external impacts like water, fire or surges but also of involuntary or voluntary deletions of files. The media should be exchangeable, e.g., CDs or magnetic tapes, and stored at a safe place. Alternatively the data can also be stored online at a remote location. Such backups are also offered as remote data transmission services (e.g., via a dedicated line).

A backup can refer to a specific application, a database, the file system or a logical or physical hard disk.

It is advisable to make a backup of the complete system before starting a new system. This also includes the operating system with all settings and user programs in the form of a so-called *image backup*, i.e., the backup of the complete hard disk. Thus, it is not necessary to reinstall the operating system and the user programs, above all the settings of the system parameters, after a hardware breakdown. The stored image can be copied directly on the hard disk; however, the dynamical data are not updated. For an update these data have to be saved at regular intervals. When the latest backup has been copied the latest changes have to be reinstalled by means of the journal file. If the journal file is no longer available the last changes can be made manually. In addition to an occasional complete backup changes are saved by incremental and differential backups to minimize the memory volume (cf. Table 7.9).

Backups are also used to file data for a later evaluation. Therefore, a filing system should be operated which not only takes technical but also organizational precautions. These are, for example, backup programs, automation of backup processes, responsibilities, filing systems and recovery scenarios. While in practice the backup often is well organized the *recovery* of damaged data is often neglected.

Packers are used to pack the data of most filing systems to minimize the data volume. Most files contain multiple characters or character sequences while text files are marked by long sequences of blanks or repeated words. But in other files, as well, many character sequences occur repeatedly. Basic methods save these multiple sequences just once and mark repetitions with a reference to their first occurrence. When sequences occur several times a repetition counter stating the multiplicity is included into the packed file. This basic principle has been extended and improved and today many different algorithms and programs can be used to pack files.

7.3 User interface

The user interface as a man–machine interface plays an important role in a WMS. If a system is accepted by the users largely depends on the right choice

Table 7.9. Backup methods in short

Type	Description	Advantages	Disadvantages
Standard	All (or all selected) files are saved and marked.	Requires just one medium or one set of media for recovery. Files can easily be found.	Backup takes a long time. Many media are required since always all files are saved.
Differential	Only changed files or files without a backup mark are saved but, unlike the incremental backup, not marked.	Only the media used for the latest standard and the latest differential backup are required for recovery. The backup takes less time than a standard backup.	The recovery generally takes longer than a standard backup.
Incremental	Only changed files and files without a backup mark are saved and marked.	Smallets memory space and quickest backup method.	The recovery takes longer than in a differential backup because of the variety of sequentially copied media.

of terminals, an intuitive operator guidance and the ergonomic design of the screen marks. This also has great effect on the error rate.

7.3.1 Terminals

Terminals are all input/output devices which – in contrast to sensors and actuators which built the interfaces to the conveyor technique – are directly operated by the user.

Table 7.10. Terminals

Typical output devices	Typical input devices

– Optical displays
  – Control lights
  – Text displays
  – Screen pages
– Printer
– Loudspeaker
	– Push buttons
	– Mouse or touch screen
	– Keyboard
	– Barcode scanner

Figure 7.15. Typical terminals with barcode scanner, text display, keyboard and radio data transmission [Photo: SYMBOL]

These input/output devices are often combined into complex units for special applications, for example:

– Mouse-controlled workstations emitting audio warnings over a loudspeaker.
– Radio terminals with text display, numerical keyboard and barcode scanner (cf. Fig. 7.15).
– Status display by a control light and acknowledgment by means of a push button.

7.3.2 Functional point of view

Terminals can be used as man-machine interface only with a suitable software and a number of functions which have to be adjusted to the respective operation. These functions interplay with the input and output and, on the one hand, should guide the user purposefully but, on the other hand, should not causelessly enforce a fixed order.

Most important, it has to be possible to cancel an already started function – as far as possible. The choice of a suitable function is supported by mostly hierarchically structured select menus. The selection of the next level and the return to the superordinate layer represent the *statical navigation* in a select menu while the return to the last chosen level represents the *dynamical navigation*. In the dynamical navigation a return can be cancelled by a forward. When this navigation is used it should be made sure that input fields which have been seized by values in the course of an earlier function are refilled with the preset values after a dynamical return.

Adequate standards have to be set for the input values and all values entered by the user have to be checked with regard to their plausibility. This means that it is not possible to enter unallowed characters, to exceed the defined value range and to infringe any interdependencies between the single parameters.

The functions carried out at a workstation or device are assigned to one or several *roles*. These functions are, for example, blocking/release of pallets and bin locations or the change of a pallet assignment. Roles are, for example, incoming goods inspections, quality checks and dispatches. In an assumed scenario the incoming goods inspection as well as the quality assurance are allowed to block pallets while only the quality assurance is entitled to release pallets. Bin locations, for example, can be blocked and released only by the dispatching.

7.3.3 Access control

Before functions are carried out in a WMS an access control should be made for safety reasons. For this purpose, the WMS administrator stores the personal names, access data and the role this person is allowed to take. Access data may be passwords or biometrical data[21]. At the start of work each person has to log in at his/her workstation and to provide the access data – e.g., by entering the password or scanning the fingerprint – for the access control. After a successful login the workstation can be used for the roles for which it has been set up and to which the respective user is entitled. The logout should be made explicitly by the user. As an alternative or additional feature, the workstation may be blocked or the user may be logged out when no entries have been made for a certain period of time. A blocked workstation can only be released by a repeated successful login of the registered user.

Workstations with an access right may also be restricted. This gives a stationary access – e.g., to exactly one or several defined I-points or order-picking stations – instead of a role-related access (storage, order-picking, etc.). Except allowed individual functions further restrictions can be made by granting certain rights (information about master data, blocking/release of batches, etc.).

The access control should offer dialogues in the user's native language (cf. section 7.3.4). The quality check may also require the traceability of users as well as of payments for services. In these access control applications personal data are recorded and stored so that at least the protection of data

[21] For example a finger print or iris image.

privacy has to be observed when such a system is introduced. It furthermore is advised to include the staff and their representatives into such projects.

7.3.4 Internationalization

This section deals with the most important aspects of the internationalization of user interfaces. Basically, inputs and outputs can be made depending on the user or the location or according to a company or international standard.

Dialogue texts can be represented in the mother tongue of the current user and can be selected manually or via the user profile of the access control. The terminals have to have the corresponding fonts or be able to load the fonts from an external device. Since the usual 8-bit code often is not sufficient to display some characters the *Unicode*, a 16-bit code was developed to support the multi-lingual word processing.

Dates and times should be displayed in a distinctive form but may also be displayed in the form common at the user's location. Decimal may be displayed in dots or the dot or comma notation but usually the regional form is chosen.

Physical values – such as weights or dimensions – should correspond to the international standard (MKS system: meter, kilogram, second). In identification suitable functions should allow for the conversion of different quantities. New programming languages support the internationalization by suitable data types and libraries.

7.3.5 Help systems and help functions

Manuals are not used during practical operations but mainly for training purposes and can also be published online (cf. Fig. 7.16) so that they can easily be searched with regard to their contents, catchwords or the complete text (full text search). More important than the online manuals, however, is a *context-sensitive help* which offers information about the current function. Of great help are displays of possible input values or value ranges. Adaptive methods can provide selective details according to the history of the user dialogue. *Adaptive help functions* are seldom used today or only in a very rudimentary form. Instead, a fixed variable, mostly a reference to the online manual, is preset which appears in case of a repeated wrong entry. Additional help functions may be reasonable in some cases:

- Electronic memo instead of "bits of paper"
- Electronic in-house mail system with a direct connection to the WMS
- Workstation-specific to-do lists for the cross-shift communication
- Workstation-specific calendar for the notification of maintenance dates and extraordinary events

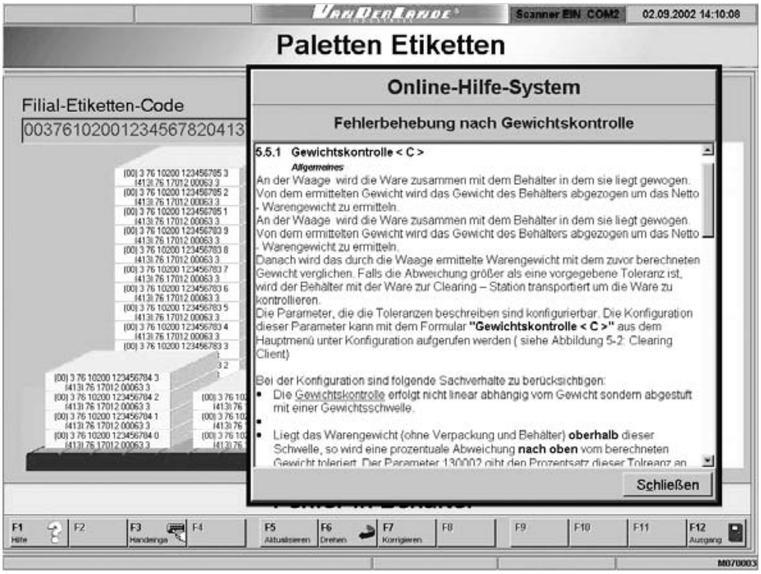


Figure 7.16. Example of an online help function [Photo: VANDERLANDE INDUSTRIES]

7.4 Operating systems

An operating system is a program of a computer system which manages and controls all components and initiates the startup of programs. It is an abstraction layer which prevents the direct access of user programs on the hardware of a computer and coordinates all activities.

Since the advent of the personal computer in the early 1980ies the term *operating system* (OS) is often used to describe characteristics of the file system, network capability and above all concepts of the user surface. An operating system, however, offers quite more elementary functions mostly hidden to the user. For users of IT systems, it makes sense to know the basic principles of an operating system which are described in short in this section. Further functions like network compatible file systems and user surfaces are described in other sections.

7.4.1 Tasks

An operating system is a software layer which protects the user programs from a direct access on the hardware. It is responsible for providing resources and services by means of the hardware. Resources are, for example, printers, barcode scanners but also the memory on a hard disk. An operating system has to perform the following tasks:

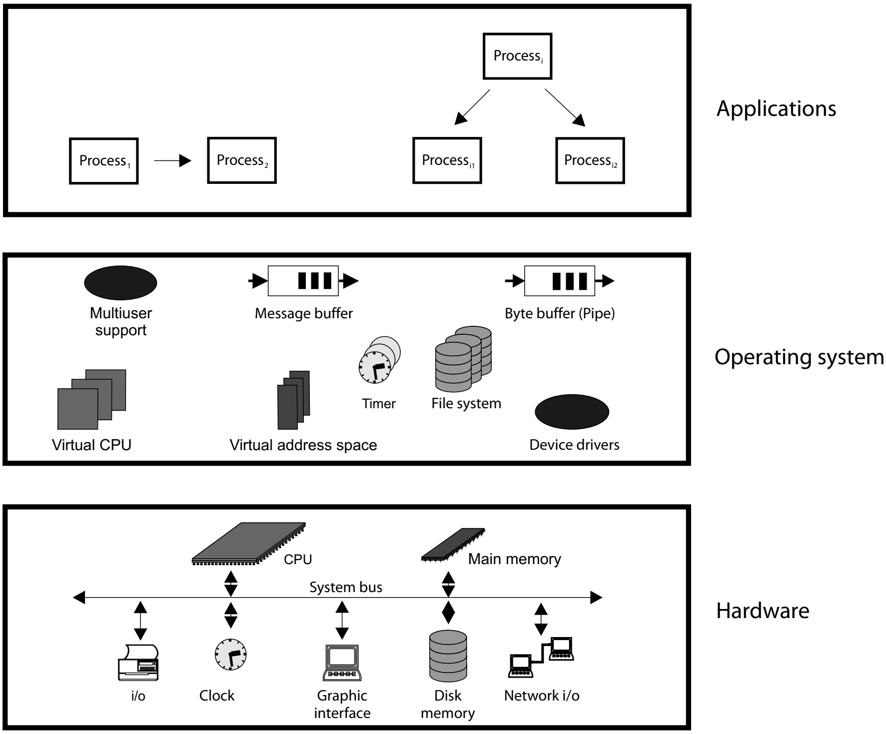


Figure 7.17. Operating systems as abstraction layer between hardware and application

- Parallel operation of several user programs (*multi-programming*).
- Implementation of defined chronological interdependencies between different user programs (*synchronization*)
- Provision of generally applicable program (*libraries*)
- Provision of a standardized input/output system (*virtual I/O system*)
- Provision of a memory management (*virtual memory*)
- Protection of the user programs against errors in other user programs
- Support of different users with user-specific rights and mutual protection (*multiuser*)

In short, the tasks of an operating system can be described as the presentation of a logical computer (*virtual machine*) with user-oriented interfaces on a high logical level.

The principle of the virtual machine as abstraction layer between hardware and user programs can still be refined further. Many modern operating systems work with a so-called hardware abstraction layer (HAL) so that a system provider can easily port his product to different hardware systems.

7.4.2 Principles

This section introduces the basic principles and components of operating systems. For further and more detailed information please refer to publications about OS [2], [76] or [101].

A basic function of each operating system is the management of different hardware systems, which are usually categorized as follows:

- Memory: The physical store where data and programs are stored during runtime.
  The hardware supports the dynamical classification into read-only and read-write areas. A *MMU* (memory management unit) permanently controls the read-only area concerning writing access which it prevents. In case of infringements it initiates an interrupt which then prompts the operating system to shut down the respective program. More advanced safety mechanisms like the access on memory areas of other programs[22] and the allocation of memory space to the programs have to be provided by the operating system (cf. p. 256). The deliberate controlled access of different programs on the same memory, however, should be supported (cf. section 7.4.2).
- Processors (CPU, Central Processing Unit): Here, the programs in the memory are run. Modern hardware supports different CPUs that have to be managed by the operating system (cf. p. 254). Modern processors have an integrated cache memory where frequently referenced commands and data are buffered to speed up the access. The efficiency of a CPU is often specified by the number of commands executable per time slot and by the clock speed. But such key values are just *one* parameter for the evaluation of the hardware performance. This is a combination of all components and without consideration of the characteristics of the operating system and the user programs these values are of no significance for the user. Instead benchmarks[23] are used to determine the performance for a group of single applications.
- Devices: These comprise the following groups of devices
  - Interfaces: Parallel and serial interfaces such as RS232 or USB (cf. Chapter 5).
  - Network: Access on local and/or long-distance networks for the communication of computers via different media (cf. section 7.1).

[22] For reasons of data protection the read-only access on memories of other programs has to be prevented in any case, cf. section 7.8. A read-write access may even be used for sabotage.

[23] In this context benchmarks are defined test runs of chosen programs of one or several application categories with defined input data running under a concrete operating system on a specified hardware. The aim is, for example, to define the response times and the medium throughput rate.

  - Bulk memory: Bulk memories with *random* access[24] such as hard disks, floppy disks, CD-ROM and DVD, which are based on rotating media and those without mechanically movable parts like flash-media cards. *One* criterion for the classification of bulk memories is the exchangeability of the storage media. This feature is of great importance for filing but also for data imports and exports and backups (cf. section 7.2.4).
  - Screen: Screens are generally controlled by so-called graphic cards, i.e., sub-systems which receive a high data rate from the CPU. Based on these data the graphic cards can independently perform complex operations. The functioning of the screen has no effect on the computer.
- Clock: The clock is a device of special importance. It initiates cyclical *interrupts* which prompt the operating system to perform certain tasks (cf. p. 254). Furthermore, the clock shows the current date and time. Operating systems offer further functions which set the time zones and, if applicable, the summer time, set the timer for processes and synchronize the computer clock in distibuted systems (cf. section 7.1.6).
  The importance of the clock was widely recognized by the public at the change of the millennium when the so-called "year-2000 problem" (Y2K problem) occurred. This problem was caused by the two-digit display of the year and mainly affected the software with the exception of computers with clocks which displayed the year in the same manner. New hardware clocks are unstructured counters which show the number of time units which have elapsed since a certain point of time.

Each user program is controlled by an operating system and allocated to a *process*. A process is also called a *task* and in addition to the user programs also includes information about the current status and the resource requirements and allocations. Tasks can be performed at any time sequentially or *concurrently*. Concurrency means that processes can be carried out parallel in time, but not necessarily. This chronological parallelism may also apply to just some time intervals, may change strictly cyclically or cover the complete runtime of a task. Since these chronological processes cannot be predicted in a dynamical system with non-deterministic external events in this context the term concurrency is used.

From the software-related point of view WMS are complex systems consisting of several, partially concurrent tasks which exchange data, provide services to each other and synchronize their program flow at defined points. Recently, the term *application* has more and more established itself for single user programs but also for more complex program systems. Accordingly, WMS are applications.

In the following the main tasks of an operating system – above all in combination with a WMS application – are described in detail.

---

[24] Random access means that a program has at any time access to any "data set" on this medium.

Control of the operating system processes The process control is responsible for performing the following tasks:

- *Mutual exclusion*: Only one process is allowed to access an exclusive resource at a time. A good example is a printer where several processes print concurrently – without further precautions. To achieve good results only one process can print at a time – the processes have to exclude themselves interactively. The principle of the mutual exclusion is also used continuously for all exclusive resources. Recent programming languages offer language constructs for the mutual exclusion.
- *Synchronization*: It has to be ensured that under certain conditions a process can continue only when it receives a signal generated by another process.
- *Avoidance of deadlocks*: It has to be avoided that a pool of processes is built where one or more processes obstructs another one[25]. Since the deadlock problem is of vital importance – not only with regard to operating systems but also for databases and material flow controls – this subject is described in more detail in section 7.4.2.
- *Communication*: The exchange of messages between processes, the so-called *interprocess communication*, has to be ensured. This communication may be limited to the exchange of signals without additional information or contain contents in the form of data sets. Furthermore, a data flow oriented communication can be used to exchange an unstructured flow of characters between the processes (cf. section 7.1).

Each process at any time represents one of the the three conditions "computing", "ready" or "blocked". Blocked processes wait for a resource, the completion of an I/O job or on a timer event. When the blocking is released the process is set to the status "ready" and applies again for a CPU. The processor management controls the allocation of processes to the processor(s) and manages the corresponding process states.

The blocked processes wait *passively*, i.e., they practically require no computing time. The operating system manages the events for which the blocked processes are waiting and "awakes" them when the event occurs. The opposite principle is the *active waiting* where a program detects an event by cyclically scanning a condition. This principle is called *polling* and is sometimes used in user programs but never in operating systems. The polling requires an idle time τ_p between the single queries so that other processes can continue with their program. If τ_p is small the polling process puts an extra load on the CPU without providing any result. If τ_p is larger this increases the average response time of $\tau_p/2$.

[25] In literature this deadlock problem is often discussed as the so-called "philosopher's problem".

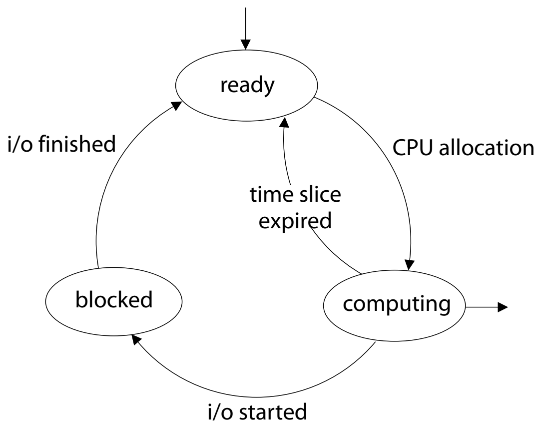


Figure 7.18. Process states

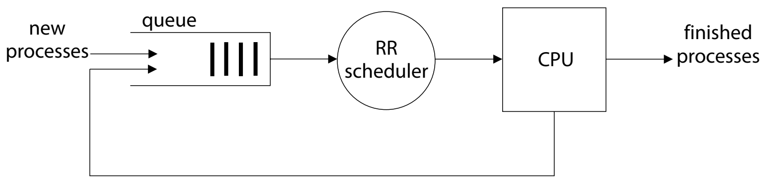


Figure 7.19. Principle of a round-robin scheduler

Processor management A ready process "applies" for the resource CPU. The corresponding allocation principle is called *scheduling* and the relevant program of an operating system is called *scheduler*. Generally, there are considerably more processes than processors. To be able to handle all processes the multiplexing is applied to the CPUs. According to this principle, the CPU is removed from a process after a certain period of time (time slice, time portion) and allocated to another process. When a scheduling is only based on the time slice and when all ready processes are allocated to a CPU strictly sequentially this is also called a round-robin scheduler (RR) (cf. Fig. 7.19).

In practice, a variety of other criteria are of importance for the scheduling. A scheduling strategy, which is implemented in almost any operating system, allocates *priorities* to the processes. These priorities are often dynamical, within certain limits: Processes which release the CPU because of an input/output get a higher priority until they reach a given upper limit. After the termination of a time slice the priority is decremented step by step up to a lower limit. The scheduler always handles the processes with the highest priority first according to the RR principles and then the processes of the next category. If a process with a higher priority is ready the current process is interrupted – i.e., it is set to the status "ready" – and the scheduler is activated.

In some operating systems the scheduler is a process itself with a high but fixed priority. Thus, processes with a priority higher than the scheduler are no longer submitted to the RR principle and can only be interrupted by a process with an even higher priority. Such processes are called *realtime processes.*

Priority-oriented scheduling methods have a dynamic which adjusts to the current system load – often according to considerably more algorithms.

Memory management Each process has a separate *virtual address space* which is provided by the operating system and represented on the physical memory by different methods. The most commonly used method makes multiple use of the physical memory space and is, therefore, also called *memory multiplexing*.

The usual procedure – the so-called *paging* – divides the memory in equal pages. Unused pages can be copied (*swapped out*) with different algorithms into a background memory, a bulk memory with slower access but a considerably larger capacity (in practice, this is the hard disk). This releases areas of the physical memory for other programs.

When swapped out pages are requested this is called a page fault and the pages are recopied (*swapped in*). At first, however, other pages have to be copied into the background to make room for the pages to be swapped in. The corresponding algorithms are very extensive and also consider the access rate on single pages as basis for a forecast [101].

The pages to be swapped out are stored in one or several files (*swapfile*, *page file*[26]) or in one or several unformatted areas of the harddisk[27] (cf. section 7.2.2). To increase the throughput by means of a concurrency many operating systems support a distributed paging on different drives.

A high page fault value per time unit always indicates a too small memory while a too small page file always leads to an error message of the operating system which is generally be indicated by warnings when a certain filling degree is exceeded. If no measures are taken the following scenarios may occur:

– The operating system suspends complete processes and swaps their memory contents in a separate background memory area, the so-called swap file.
– the operating system by force closes down single processes.
– The problem is ignored what leads to a deadlock (cf. section 7.4.2). Such a deadlock can often not be cancelled by closing a process because this would require additional memory space which is no longer available.

[26] In some systems the term swap file is used while in others this term is not used for pagings but for complete swappings.
[27] Hard disks are mostly divided into *partitions* which may contain file systems or be used as *raw device* without a formatted applications or databases.

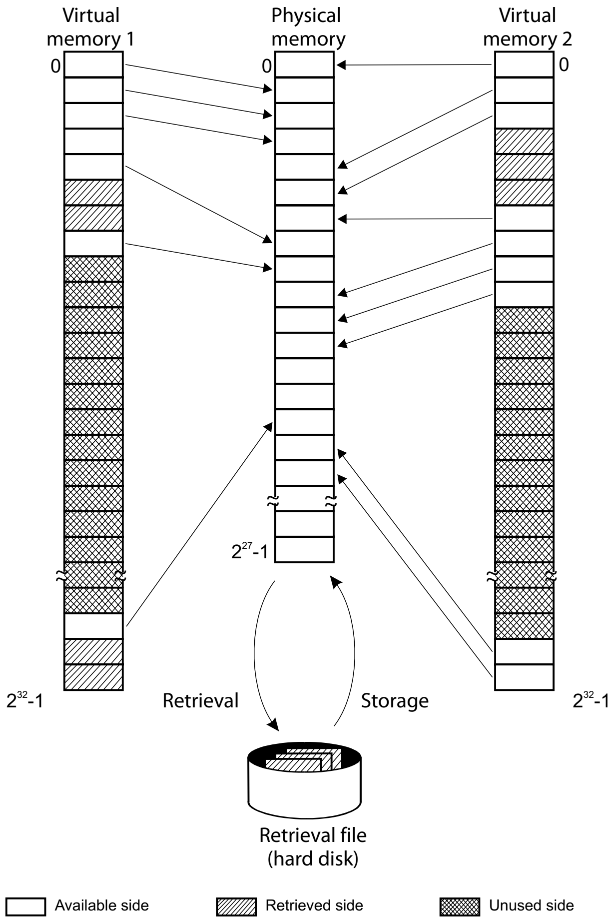


Figure 7.20. Representation of two virtual address spaces on a physical memory at the example of a 32-bit computer with 128 MBytes physical main memory

Device and resource management On its bottom level the device management communicates in realtime with the hardware. This is done by *interrupts* prompted by the hardware which interrupt the program flow and branch into an *interrupt service routine*. This routine handles the interrupts by exchanging data and signals with the requested device and then passes the control back to the up to now idle program flow.

On the next higher logical level the devices are managed as *resources* or as *operating facilities*. A resource is not necessarily a physical device. Logical or virtual resources are, for example, are communication buffers, memory areas used by several processes or database tables. They are generally used *exclusively* by processes in case of a write access. Read-only accesses gen-

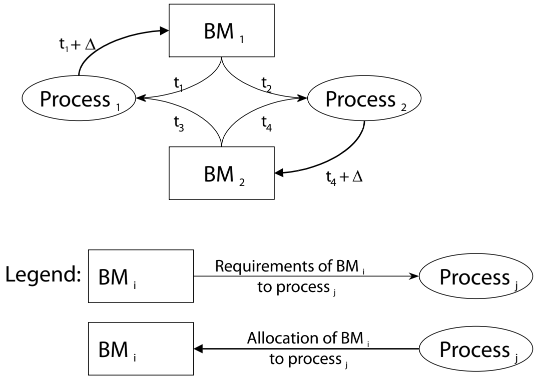


Figure 7.21. Deadlock situation in a resource allocation graph

erally lead to no status changes and can thus be performed concurrently. This method is called *mutual exclusion* or in short *mutex*. The period of time during which a resource is exclusively occupied by a process is also called a *critical region*. Processes which want to enter a critical region at first have to request the respective resource from the operating system. If this is available it is allocated; if not the process awaits its availability. The resource management awakes the process as soon as the corresponding resource is available. If several processes wait for the same resource their requests are managed in a waiting queue according to the first-come-first-serve principle (*FCFS*)[28] In case of time-critical processes or for processes which occupy some other resources it is possible and sometimes reasonable to diverge from the strict FCFS principle.

The deadlock problem In an environments where the mutual exclusion is ensured, processes wait infinitely for the availability of resources and where it is impossible to remove a resource from a process deadlock problems may occur. A deadlock is a system blockade which can only be released by the removal of resources *preemption* or by the enforced shut down of processes.

Figure 7.21 shows a typical deadlock situation by means of a *resource allocation graph*. Processes and resources build the node of such a graph while resource requests and allocations are represented by the edges. The graph describes the following situation over the time for two processes (P_1, P_2) and two resources: (BM_1, BM_2):

[28] According to the first-come-first-serve principle the first process to request a resource is the first to be allocated to a resource, similar to the FIFO principle which is also common in material flow (section 2.2.3).

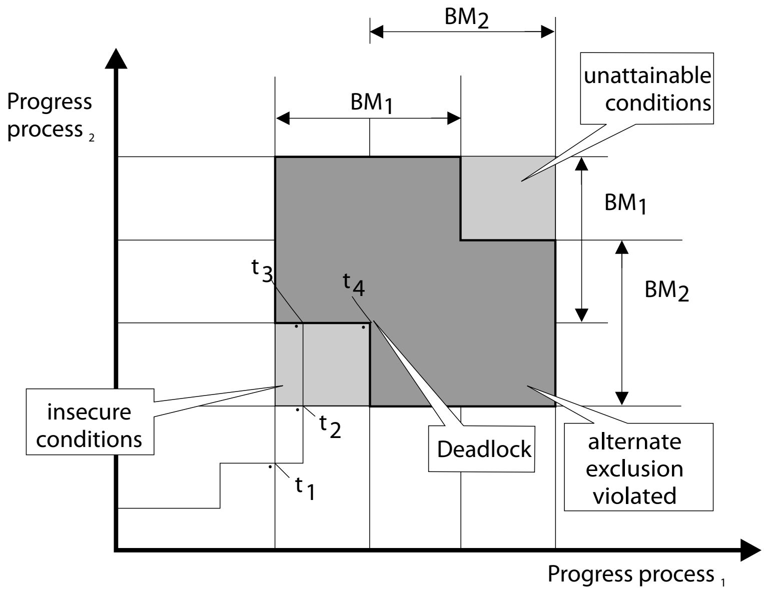


Figure 7.22. Deadlock situation in a process advancement diagram

$$\begin{aligned}
t_1 &: P_1 \text{ requests } BM_1\\
t_1 + \Delta &: BM_1 \text{ is allocated to } P_1\\
t_2 &: P_2 \text{ requests } BM_2\\
t_2 + \Delta &: BM_2 \text{ is allocated to } P_2\\
t_3 &: P_1 \text{ requests } BM_2 \text{ and has to wait}\\
t_4 &: P_2 \text{ requests } BM_1 \text{ and has to wait}
\end{aligned}$$

At the time t_4 a deadlock occurs. Neither of the two processes can continue without having to wait for the other.

Another way to represent the above-mentioned scenario is a diagram of the process advancements (cf. Fig. 7.22). It shows by means of a trajectory how such situations occur and how they can be *avoided*. Both axes of the coordinate system represent the advancement of the two processes P_1 and P_2. The graphs in such a diagram represent the advancement of both processes and may be marked with the parameter t for time. Such graphs are also called *trajectories*.

Under the assumption that a preemption is not allowed, these graphs never contain a component in the negative axis direction. In case of a one-processor system they always build ascending step curves. The resource required by both processes is entered at the axes. To ensure the mutual exclusion must never cross a dark grey area. This would inevitably lead to a deadlock when the area at the lower left side – the *unsafe conditions* – is reached. Here, the *deadlock prevention* methods come into force which are used by operating systems and sometimes also by databases.

An alternative to the deadlock prevention is to request all required resources by an indivisible operation or to solve a deadlock problem by preemption. If all required resources are allocated on the first request of just one resource this generally reduces the throughput because other processes have to wait for this resource. But a deadlock can be avoided with this method. After a preemption the operations performed up to then by the respective resources become invalid and have to be repeated later on. This also leads to a reduced throughput.

These interrelations also apply to material flow systems. In Chapter 9 a warehouse is described which is also prone to deadlock problems which have to be avoided by means of suitable algorithms. A preemption is not advisable and not always possible in this case because returns would be necessary after transport interrupts.

Clocks The hardware clock sets the basis for

– the initiation of cyclical functions of the operating system such as the scheduling which is hidden to the user
– the provision of timers
– the calculation of the current date and time

The hardware clock of a computer is always subjected to a certain drift so that it has to be updated from time to time. This may be done manually by a system administrator, by coupling it to an external reference clock ro by means of a clock protocol (cf. section 7.1.6) in a computer network. The clock must not be reset while the system is running because this would interfere with the causality. If absolute times are used in a system a clock reset may imply that the effect of an uninvolved process lies before its cause. For this reason, hardware clocks deliberately lose a little bit.

Such an effect occurs, for example, when applications process files according to the date of their creation or change. In software development all programs are interpreted by the "make" command the files of which have been created or changed before the last successful "make" request. If the clock had been reset during the development, the program may be interpreted only partially what leads to incompatible results and thus to programs which operate faultily or not at all. Similar effects may also occur in WMS when dates are used as criteria for the execution of single work steps.

However, if a clock has to be reset this problem can be solved by means of a *software clock*. This is always based on a hardware clock but in addition to this offers an offset variable for time corrections. The software clock misses single ticks so that the offset is increased and the clock runs slower but never backwards. In network systems this problem is solved by means of clock protocols.

7.5 Programming languages

Programming languages are used to compile logical concepts into functioning programs. They are a link between the program P of a programmer and the machine M where the program should run. The term "machine" should not be mixed up with the hardware because the operating system is a virtual machine for the execution of programs[29]. Each program processes inputs I and generates outputs O which in turn may be inputs for another program.

7.5.1 Compilers and interpreters

This section deals with the most important concepts and defines some terms. The programming language of a program should be of minor importance to the user as long as the provider guarantees for the correct functioning and is available in case of changes, extensions and portings to another hardware. But if this is not the case or when other providers should or have to be involved the programming language becomes more important.

T diagrams (cf. Fig. 7.23) are a kind of graphical representation for these interrelations [105]. During the runtime each program needs a machine which is able to perform the program code. Seldom machines are available which can directly handle the program code. This problem can be solved in two ways:

- By means of a *compiler* the program code is interpreted into the language which can be read by the machine. The program to be interpreted is called *source program* or *source code*, the result of such a compilation is the *machine code*, which can be performed by a machine. Compilers can be divided into the following two types:
  - *Assemblers* are structure conserving, i.e., the program language can be read by man but can also be performed by a machine. In this case, the programs are written in an assembler language and compiled by an assembler into the machine language. This kind of language is no longer used for the programming of applications but for hardware-oriented developments.
  - *Compiler* are interpreters which do not maintain the structure of the program to be interpreted (cf. Fig. 7.25). Owing to this transformation the compiler can perform extensive optimizations with the aim to create small programs which require less memory space and short execution times. Since the structure is transformed the compiling cannot be reverted so that the source code has always to be available for further changes of the program logic. Most providers of WMS do not supply the source code or just parts of it while the complete source code is provided in *open-source projects.*

[29] The operating system is a program as well but it hides the hardware for the applications discussed here.

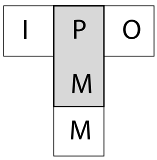


Figure 7.23. T-diagram. A program P runs on a machine M. It reads input data I and generates output data O

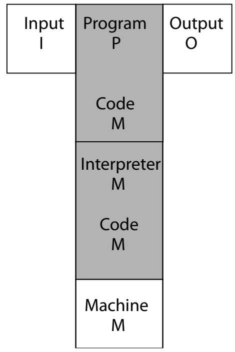


Figure 7.24. Interpreter as virtual machine. An interpreter runs on a host M_1 and runs a program P written in the language M

  - *Macro-expanders* or *Macro-processors* are text replacement systems where required program parts are replaced by parameterizable text modules, the so-called *macros*. A macro-processor replaces the formal parameters of the macros by the current parameters of the macro request. Applications are not programmed by macros but some programming languages use macros. The actual compiling is preceded by a separate *pass*[30] where the macros are expanded. On the one hand, this reduces the desk work but, on the other hand, hides weak points in the programming language [75].
- The source code is directly performed with the support of an *interpreter*, i.e., a program which runs on a *host* and interprets and performs the commands of an application (cf. Fig. 7.24). Thus, an interpreter is a *virtual machine* which hides the features of the host to the applications.

Applications are developed with compilers and/or interpreters. Both basic principles have advantages as well as disadvantages:

- An interpreter facilitates the immediate performance of a program without an interpretation. When an interpreter is used the execution times are longer than for compiled programs.
- Compilers analyze the program during the *compile time* while interpreters analyze it during the *runtime*, i.e., when the program is being executed. This analysis is able to identify programs with a faulty syntax. When it is

[30] A pass is the performance of the complete source code by a program. Compilers, for example, operate with several consecutive passes such as syntax analysis, code generation and code optimization.

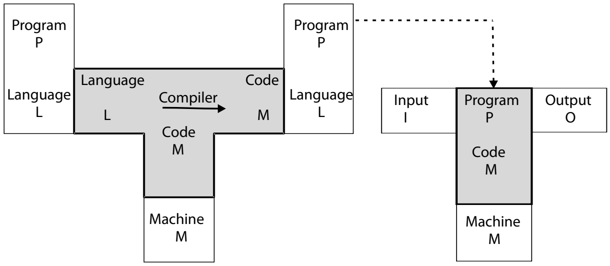


Figure 7.25. Compilation of a program P written in the source code S into the target code M. The compiler runs on the machine M. The resulting target code then runs on a machine M.

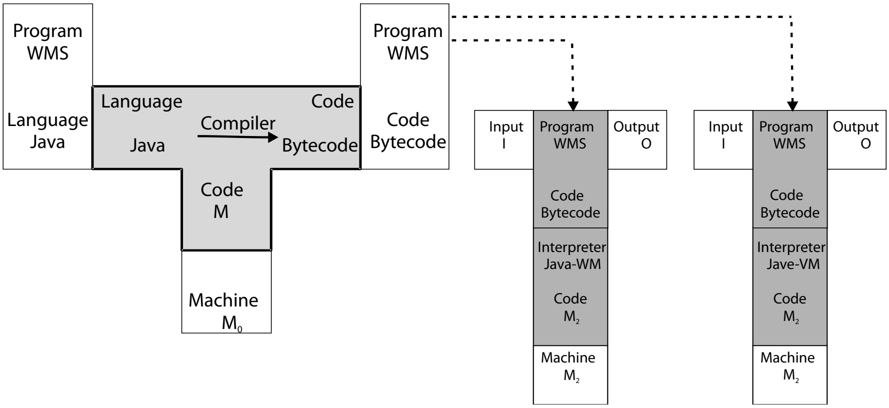


Figure 7.26. Interpretation of a WMS in JAVA running on the machine M in byte code. This portable byte code can be performed by an interpreter (virtual machine) optionally on machine M_1 or M_2.

performed only during the runtime one of its potentials is not completely utilized.

– Compiled programs are not portable. Thus, an application can only run on the computer and under the operating systems it was designed for. In some cases, it is even restricted to the version of the operating system. To be able to run the program on another computer at least the source code and a suitable compiler have to be available while a program can run on any computer with a corresponding interpreter.

A synthesis of the compilation and interpretation principles offers a good compromise between execution speed, compile time analysis and portability. The source code is compiled into a machine-independent intermediate code which can be performed by all computers with an adequate interpreter.

The most commonly used programming system which is based on this concept is Java. A source code written in Java is interpreted by a Java compiler into the so-called *bytecode*, a binary code which is independent of the hardware and an operating system (cf. Fig. 7.26). Thus, all lexical, syntax and some semantical errors are identified already during the compiling (cf. section 7.5.2). In addition to this, the compiler can optimize the code. The resulting byte code is very compact and is interpreted by a virtual Java machine (*VM*) during the runtime. This concept is suitable for the creation of distributed applications. kap9 describes a Java-based WMS.

7.5.2 Language concepts

Programming languages are described by

– their lexical elements
– their syntax structure
– their semantics

The lexical elements are the atomic units, the "characters" of the source code, i.e., keywords, numerical constants, character sequences with a fixed contents, operators and identifiers for variables and functions.

The syntax structure describes the "sentence structure" of a program and is formally described by a *recursive definition*, by the *Backus–Naur form* (*BNF*) or a similar notation or by *Syntax diagrams.* Syntaxes are represented by *grammar symbols.* The syntax of a programming language is specified by a number of *rules* with exactly one grammar symbol on the left side and a sequence of grammar symbols and lexical symbols on the right side. Syntax diagrams graphically represent these rules by rectangles and circles or rounded rectangles. In contrast to the syntax diagrams the BNF shows special operators and symbols on its right side to simplify the representation. Optional symbols or symbol sequences are put into squared brackets while accolades indicate that the corresponding symbol sequence is iterative and may thus appear in a program in series (cf. Fig. 7.27).

The semantics of a programming language generally is described verbally to the programmers. One example of semantics is the observance of signatures, i.e., when a function is defined and requested the number and types of parameters have to correspond to the declaration.

7.5.3 Language generations

Programming languages can be categorized as follows:

1. First generation: This is the machine code, i.e., the language which can be directly performed by the computer hardware without interpretation.
2. Second generation: The assembler code has the same structure as the languages of the first generation and facilitates the programming by means of mnemonic codes, i.e., plain writing identifiers for commands and storage addresses. Before this code is performed it always has to be interpreted by an assembler.

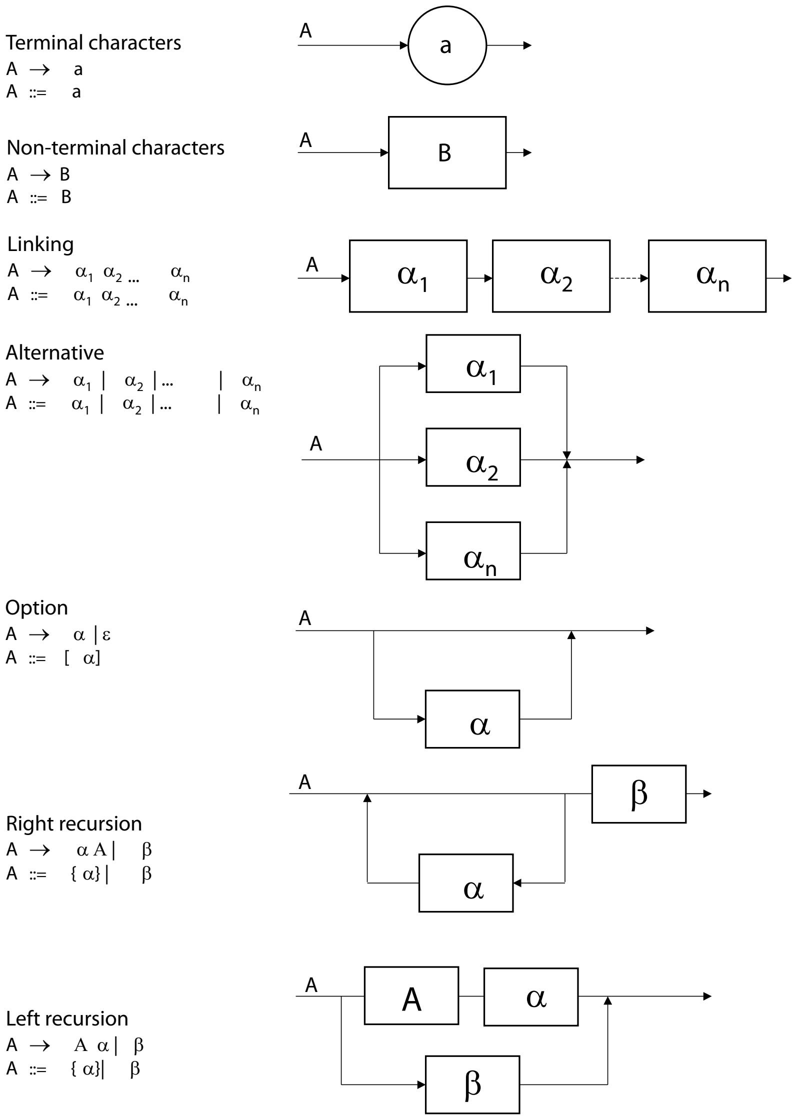


Figure 7.27. Representation of the syntax of a programming language by recursive definitions, by the Backus–Naur form and by syntax diagrams

3. Third generation: The problem-oriented programming languages are often considered as advanced programming languages and can be interpreted either directly or after a compiling – depending on the language type and the availability of the interpreter and compiler.
4. Fourth generation: (*4GL*: 4^{th} Generation Language): Programs are mainly written in natural language. The necessary data structures and algorithms are generated or chosen by a compiler and/or interpreter. A typical example of a 4GL language is SGL which is used for database queries and manipulations.

7.6 Basic principles of object-oriented programming

This section describes the basic idea of the object-oriented programming. Object-orientation is not just a programming technique but also a method to analyze and design systems. The last sub-section shortly deals with these aspects. Transport devices are described which are able to transport unit loads from one location to another. These images include UML symbols (*unified modelling language*, (cf. section 7.6.4) as well as code pieces written in Java.

7.6.1 Data abstraction

The traditional programming is either based on data structures or the functions. The basic idea of data abstraction is to combine data and functions into a unit. These data, which may consist of such units, are accessed *exclusively* via these functions. This principle is also called information hiding and prevents a direct access on the data[31] and offers the following advantages:

– The standardized interfaces of the function query are abstracted from the basic data
– The details of the access functions and the underlying data are hidden to the initiator.
– The data integrity is assured exclusively by the access functions and not by the initiator
– Additional changes can be made locally and without affecting their initiators.

At first this exclusive use of the function interface and the explicit prohibition of a direct access seems very complex for "simple" variables. The advantages become visible at a second sight when the software is upgraded and when at each access on a variable other variables have to be changed or other functions have to be requested. Only the corresponding functions

[31] However, many programming languages which support this concept grant direct data access. This has to be prevented by programming disciplines.

have to be upgraded while the interface to the program remains unchanged. To use this function the programmer just needs a self-explanatory name; the details are hidden to him.

For example, the additional request can be made in a WMS to record the time for which a storage bin has been blocked in an operating protocol. In Java the change could directly be programmed as follows[32]:

```
...
// Choice of a storage bin bin
...
// The chosen bin is marked as blocked
bin.blocked = true;
...
```

In this programming language all allocations to the variable `block` have to be replaced later on by corresponding function requests or be completed by additional commands. When the information hiding principle is used this line `bin.blocked = true` has to be replaced by `bin.block()`, i.e., corresponding function requests. The function `block` allocated to the bin would be programmed as follows:

```
class bin {
  ...
  public void block()   { blocked = true;  }
  public void unblock() { blocked = false; }
  ...
}
```

By extending the data type bin by the variable `startBlock` the functions can now be extended as follows:

```
class bin {
  long startBlock;
  ...
  public void block() {
    if (! blocked) {
      block = true;
      startBlock = getTime_ms();
    }
  }
  public void unblock() {
    long delta;
    if (blocked) {
      block = false;
      delta = (System.currentTimeMillis() - startBlock)/1000/60;
      logbook.println( "bin " +  name +
                       "was blocked" + delta +
                       "minutes.");
    }
  }
  ...
}
```

[32] The double slash marks the rest of the respective line as a comment.

In this case, the request position remains the same so that the number of software parts to be changed and tested remains small.

7.6.2 Classes and objects

The *description* of the entity of all logically related data descriptions and functions is called *class*, the data are called *attributes* and the functions *methods* of this class. A class *contains* no attribute *values*[33] but only meta-data. To generate data at first an *object* of this class has to be generated which is also called the *instance* of its class and has a memory for the storage of attribute values. Any number of objects can be generated from each class under consideration of the limited memory space. In most programming languages the operator for the instantiation of classes is set to `new`.

This principle, however, does not guarantee that all attributes are initialized. For this purpose, the object-based technique offers special functions, the *constructors*. In most programming languages constructors have to bear the name of their class and each class has to contain at least one constructor[34]. If needed, several constructors can be specified when they have different parameter lists (signatures)[35]. The initialization of a class at first provides memory space for its attributes and then carries out the constructor requested by the programmer.

A rack feeder, for example, may be represented by a class AS/RS. This class may contain several constructors which are shown below in a very simple Java code:

```
class AS/RS {
  // Definition of a constant for the idle position of all AS/RS
  private final position homepos = new  Position(1, 1);

  // Attribute for the current position of the AS/RS
  private Position pos;

  AS/RS () {
    // Generation of a new instance for the attribute pos
    pos = new Position();
    // Setting the position coordinates from the constant
    // homepos
    pos.setx(homepos.x());
    pos.sety(homepos.y());
  }

  AS/RS (int x, int y) {
    // Generation of a new instance for the attribute pos
```

[33] The *static* attributes to which values of a class are allocated are not studied here.

[34] Most programming languages provide implicit default constructors which, however, cannot replace a specific initialization of attributes with the interdependencies of their values.

[35] This overload technique is also allowed for methods.

```
    pos = new Position();
    // Setting the position coordinates which are transferred
    // to this constructor as parameter
    pos.setx(x);
    pos.sety(y);
  }

  S/RS (String pos) {
    // Generation of a new instance for the attribute pos
    pos = new Position();
    // The class position can calculate the coordinate values
    // also from a string
    pos.setByName(s);
  }
}
```

A new instance of a rack feeder can be generated in three different ways:

– `AS/RS as/rs1 = new AS/RS()` for a new rack feeder. The newly created object of the class AS/RS is named as/rs1 and is created with the standard initialization, the idle position (1,).
– `AS/RS as/rs1 = new AS/RS(2, 3)` for a new rack feeder with the current position (2,3).
– `AS/RS as/rs = new AS/RS("E2T3")` for a new rack feeder with current position which is described by the character sequence E2T3.

The opposite of a constructor is the *destructor* which releases the resources occupied by the object. Finally, the destructor releases the memory space occupied by the object[36].

7.6.3 Inheritance

An analysis of objects represented by a software shows that these have joint functions or attributes. A "conveyor", for example, can transport unit loads from one storage location to another. This can be achieved by different technical features which often offer additional functions. The basic idea of the object *orientation* adds the specialization of classes to the above-mentioned *object-based* concept .

A class which provides methods and attributes can be specialized by a *derivation*. The more general class is called *basic class*, the more specialized class *derived class*. This principle is also called *inheritance* since the derived classes "inherit" the attributes and methods of their basic class.

The principle of inheritance is shown in Fig. 7.28 at the example of the basic class `transport facility` for all transport facilities and a derivation

[36] Because of the automatic memory management in Java systems it cannot be predicted when the destructor is performed. In C++ systems destructors are performed immediately when a delete operator is used for an object. The concepts differ considerably but this does not impair basic understanding of this technique.

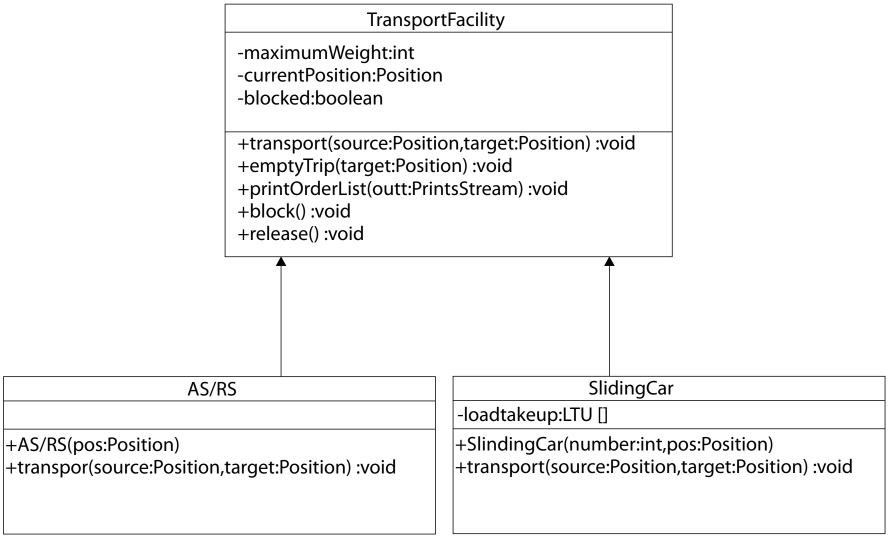


Figure 7.28. Derivation of one class each of `AS/RS` and `sliding cars` from the basic class `transport facilities`. Attributes are marked with "-" and methods with "+"

each for rack feeders and sliding cars. Both derived classes can be instantiated while in this case no objects can be generated from the basic class. The basic class just provides some attributes and *declares* some methods, i.e., it specifies an interface. WMS which include AS/RS as well as sliding cars can operate with the methods of the class `transport facility`. Only during the instantiation the programmer has to take care that the object is generated *exactly* from the class which describes this transport facility. The programming system ensures that always the correct method is called.

If methods are defined by a basic class as well as by the derived class the methods of the derivation are requested; otherwise the methods of the basic class.

Derivated classes can include further attributes and methods which are not declared in the basic class. Such methods cannot be requested via the interface of the basic class because they are unknown there[37]. One example is a curve-going AS/RS (allows to convert the AS/RS in different aisles) which can swap the aisle by means of the method `changeAisle()`.

The derivation principle can be used in several ways so that a complete tree of other classes can be derived from a basic class, e.g., the class AS/RS.

[37] In this case the object has to be converted into an object of the derived type (*type casting*).

7.6.4 Unified modelling language

The object-oriented *programming* is just one aspect of the object-oriented software which includes a comprehensive approach from analyses over the design and programming up to tests. For this purpose, a variety of software development tools has been developed, e.g., the unified modelling language (*UML*) to simplify the notations. This modelling language includes a number of graphical notations for very different views; the most common diagrams are:

- A use case diagram describes the interaction of actuators – in this case the WMS operator – with the system to be analyzed. This kind of diagram is used to informally represent such interrelations.
- Class diagrams represent the derivation relationship and associations between classes. Fig. 7.28 shows a derivation relationship in a UML notation while associations are shown in Fig. 7.13 similar to associations in an entity relationship diagram.
- Interaction diagrams describe the collaboration behavior of object groups: sequence and collaboration diagrams. Sequence diagrams (cf. Fig. 7.29) show the behavior of objects along their "lifeline" which can be seen as a kind of time axis[38]. Collaboration diagrams offer the same information without considering the lifeline.
- Activity diagrams (cf. Fig. 7.30) represent a series of activities, either conditional or concurrent.

The system described in Chapter 9 was designed with UML. The diagrams can be found on the Internet under `www.mywms.org`.

7.7 Extensible markup language: XML

XML is a meta language for the free definition of other languages by which data structures can easily be described, corresponding values can be specified and transported to the data. Owing to the suggestive structure of XML and the variety of tools which are available in different versions and qualities data flows can easily be processed in the XML format. This ensures a high compatibility between different software and hardware products.

7.7.1 Key-value-coding

In the past data were often transported by *telegrams* where the values have to be written consecutively and in a given order. Unlike the XML format in

[38] Here, the main focus is not set on time but cause-effect relationships are studied. This difference is important above all for distributed applications.

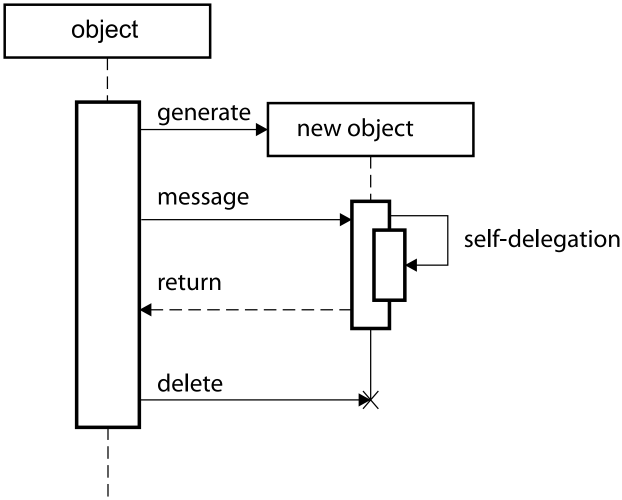


Figure 7.29. Example of a UML sequence diagram

a telegram the data are placed in a given order on fixed positions or they are separated by delimiters. To be able to plan such telegrams the data have to be known in detail because to change them afterward would require considerable effort. If not all data are sent within a telegram the corresponding positions in the related *byte stream* have to be filled with zeroes or marked as empty.

XML uses the so-called key-value coding which analogous to a dictionary represents one or several values for each key. Thus, there never is a value without a key. On the other hand, there may be a key without a value which may even be reasonable. As an extension of the pure key-value coding XML allows for the multiple denotation of a key.

The key values of an XML document are combined into elements where the key is written as *element name* and the value as *element content* within the *element tag*[39]. Written in angle brackets the element name builds the *start tag* introducing the element. It is followed by the element content and then by the end tag which, also in angle brackets and preceded by a slash repeats the element name.

The contents of elements may consist of further elements. If, for example, a stored article, e.g., a can of peas, should be described in XML this could look as follows whereas the characteristics of the article are framed by the start tag `<article>` and the end tag `</article>`:

```
<article>
  <ean>401234500001</ean>
  <name>peas</name>
  <quantity>850</quantity>
```

[39] A tag is the character sequence which gives the element its name.

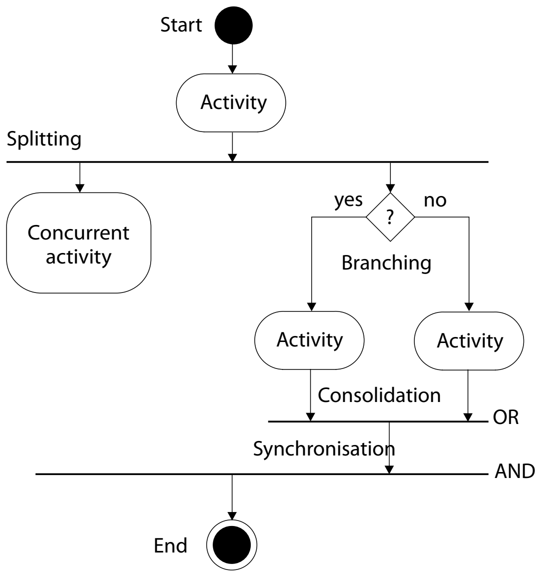


Figure 7.30. Example of a UML activity diagram

```
    <unit>ml</unit>
    <packaging>can</packaging>
</article>
```

Alternatively, attributes can be used in XML which contain the characteristics of an element. The attribute values have to be included in pairs in apostrophes or quotations. If an element has no content, the end tag can be omitted if the last character before the closing angle bracket is a slash. This is no longer called a start and end tag but an *empty element tag*. The article is then described as follows:

```
<article
    ean="401234500001"
    name="peas"
    quantity="850"
    unit="ml"
    packaging="can"
/>
```

If mnemonic (self-explaining) names are used to denote the elements, XML can be used to get a self-documenting description of data structures. The above example of an article element shows that it is possible to use *sub-elements* and to build attributes for elements. It is also possible to mix both procedures as is shown below at a customer description, which, because of the self-explaining element and attribute names, has not been explained further:

```
<customer title="Mr" name="Duck" surname="Donald"
           customer no="123">
  <addresse type="supplier" road="Minnie Street 3"
           city="Ducktown" zipcode"44227"/>
  <addresse type="invoice" road="Goose Road 42"
           city="Ducktown" zipcode="44227"/>
  <telephone type="work">+49-231-21282</telephone>
  <telephone type="private">+49-231-56080</telephone>
</customer>
```

The resulting XML documents, i.e., documents which have a start tag and an end tag for an element (in the correct order and nesting) or just an element empty tag and where the attributes of the start tag and the element empty tag are used as shown above are called well-formed or admissible. Below, an example of an inadmissible XML document:

```
<tag1>
  <tag2>
  </tag1>
</tag2>
<tag3>
  <tag4>
</tag3>
```

Generally it can be said that XML, if correctly used, in combination with the key-value coding has many advantages compared to the telegram. In addition to self-documentation, attention should also be given to the fact that there is no defined order of elements (and of their attributes). Another advantage is that optional fields can simply be omitted.

7.7.2 The syntax of XML

With the help of DTD, the *document type definitions* of schemes XML documents can be restricted. Such a restriction is important to *validate* this document. While the admissibility of an XML document can be derived from its structure, i.e., the correct handling of elements independent of its contents the scheme or the DTD specifies the validity of a document by checking the existence or absence of certain elements as well as the observance of nestings.

The example below shows a DTD for the above example of a customer element. The element `customer` consists at least of one element `address` and any number of `telephone` elements whereas `address` consists of an element empty tag and `telephone` of a start and an end tag:

```
<!ELEMENT customer (address+,telephone*)>
<!ELEMENT address EMPTY>
<!ELEMENT telephone (#PCDATA)>
<!ATTLIST customer
  title          CDATA #REQUIRED
  name           CDATA #REQUIRED
  surname        CDATA #IMPLIED
  customer no    CDATA #IMPLIED  >
<!ATTLIST addresse
  type           (supplier|invoice) "invoice"
  road           CDATA #REQUIRED
  city           CDATA #REQUIRED
  zipcode        CDATA #REQUIRED >
<!ATTLIST telephone
  typ e          (work|mobile|private) "private" >
```

DTDs should be used with care because in the element above `telephone` with the attribute `type=`“`private`“ may occur several times in a customer element or `telephone` may consist of an element empty tag and contain no number at all. In the last case, the telephone number could be defined as an attribute in `telephone`. However, DTD cannot check the correctness of a telephone number; any entered character sequence would be valid.

The content of XML documents can be checked much better by means of schemes which are able to analyze also attributes and element contents.

7.7.3 Parsers and processors

The basic tool for the processing of XML documents are parsers and processors, although with different goals: While the processor converts an existing XML document into another form using a *style sheet* the parser reads the document and provides the included elements and their attributes for further processing.

There are two types of parsers: On the one hand, the DOM, the *Document Object Model* and, on the other hand, the SAX – *simple API for XML*. The difference between the two parser models is the way how they process the XML documents. While DOM reads the complete document and represents it in the form of a DOM tree[40] with a root element SAX operates according to the *stream editor* (comparable to sed under UNIX) and immediately returns each completely recognized construction.

DOM and SAX behave differently: While DOM works more slowly because of its tree structure SAX processes, the elements and thus the complete

[40] In computer sciences a tree is often used to represent nested structures.

document relatively quick. If SAX detects an error in a long XML document the actions are difficult to reverse. DOM, on the other hand, starts the processing only after the complete tree has been built from an XML document and the document has been validated.

However, DOM needs much more memory space and in case of very large XML documents it may happen that it cannot be processed with DOM. Unlike the SAX concept the DOM parser is able to generate XML documents and to process existing documents interactively.

The different parsers are provided by the developers partly as freeware[41]. Parsers often have a different performance. Parsers and processors are available with standardized interfaces for different programming languages (e.g., C, C++ and Java) and can directly be integrated into existing applications.

7.7.4 Variety with style sheets

The example of `customer` shown in section 7.7.1 could be converted into an order by adding an element `orderItem` to a new element `order`. The element `orderItem` is based on the element `article` shown in section 7.7.1 to which it adds `price` and `number`. The respective DTD looks as follows:

```
<!ELEMENT order (customer,orderItem+)>
<!ELEMENT customer       ...             >
<!ATTLIST customer       ...             >
<!ELEMENT orderItem EMPTY>
<!ATTLIST orderItem ean        CDATA  #REQUIRED
                    name       CDATA  #REQUIRED
                    quantity   CDATA  #REQUIRED
                    unit       CDATA  #REQUIRED
                    packaging  CDATA  #REQUIRED
                    price      CDATA  #REQUIRED
                    number     CDATA  #REQUIRED >
```

With the DTD above XML documents of the type `order` can be validated. The following example shows a well-formed and admissible XML document:

```
<order>
  <customer titel="Mrs" name="Marple"  surname="Jane"
         customer no="0042">
    <addresse  type="invoice" road="Highstreet 42"
              city="St. Marymead"      zipcode="1650" />
  </customer>
  <orderItem  ean="401234500001"       name="peas"
              packaging="can"         unit="ml"
              quantity="850"  price="0,74"  number="03" />
</order>
```

[41] E.g., under http://xml.apache.org.

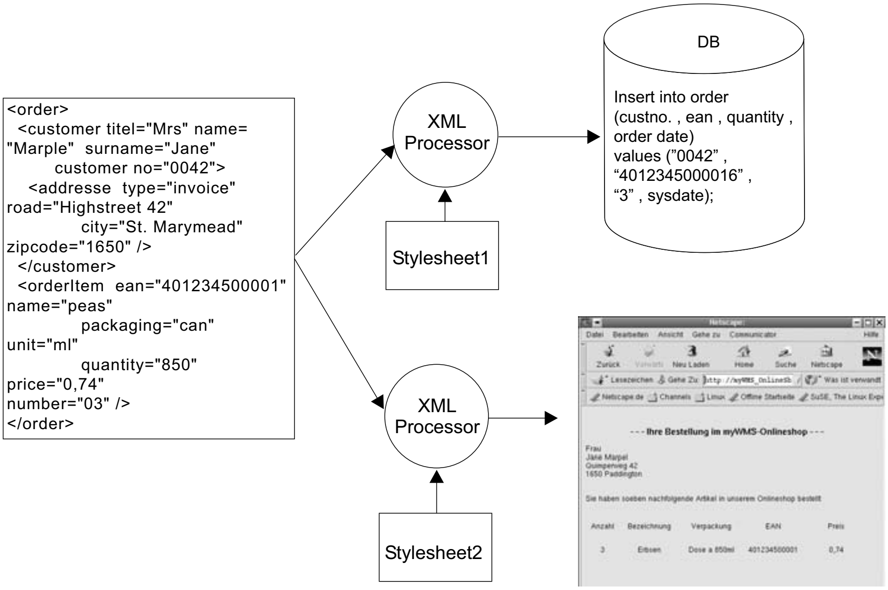


Figure 7.31. Conversion of an XML document by means of a style sheet

An order written in XML can be converted with a processor and a suitable style sheet. The output format is determined by the contents of the style sheet so that different results can be achieved with the same processor and input data but different style sheets.

Figure 7.31 shows two differently formatted orders.While the processor converts the XML document with a style sheet into an SQL statement for database processing the same processor (or another one) converts the XML document into a browser compatible HTML file.

7.8 Safety aspects

When data worth being protected are exchanged via networks safety regulations have to be set up with the aim of ensuring the confidentiality, integrity and assignment of data to their owner or sender as well as their availability. These safety regulations have to protect the data against random events as well as against targeted attacks. Cryptography is an important tool for ensuring the confidentiality, integrity and assignment of data.

The availability of data is achieved by diversity and distribution – both can be ensured by backups (cf.. section 7.2.4) – as well as by error tolerances [99].

In addition to the contents of a communication its conditions may also be worth to be protected. The most important aspects are the anonymity and protection of the communication partners. These may be persons, programs or processes of the operating system. A communication profile provides rough data based on which conclusions can be drawn concerning warehouse movements, customer behavior or sales volumes with no or limited knowledge of the communication contents.

Physically distributed WMS and/or WMS in network environments are particularly vulnerable. E-commerce systems need billing processes where the quality characteristic *safety* plays a major role. In the following sections the principles are described in the example of an e-mail exchange which can easily be transferred to other scenarios in distributed applications.

7.8.1 Secrecy

Messages have been kept secret on their way from sender to recipient for several thousand years. In recent years new procedures have been developed which also meet highest safety requirements.

To keep data secret or otherwise safe they are encoded. The basic encoding principle is based on an encryption as well as a decryption algorithm using keys (cf. Fig. 7.32). These algorithms and keys can be structured quite differently. The safety of the code, however, should always be ensured by the key and not by the algorithm. By means of the key K the message M is compiled in an encoded message M_K with the encoding function E[42].

$$E(K, M) \rightarrow M_K$$

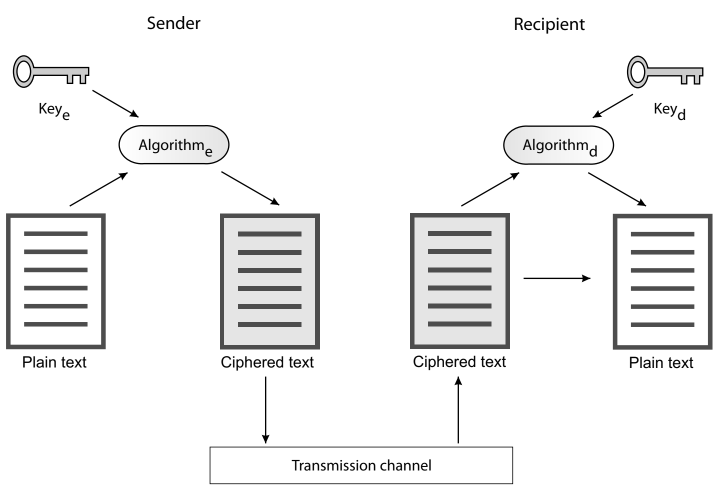


Figure 7.32. Basic encoding principle

[42] E stands for *encryption*.

Symmetrical or *secret* encoding methods use the same key for encryption and decryption. The main problem here is the secure exchange of keys. For this purpose, sender and recipient need a separate transmission channel because the channel used for the transmission of messages is not encrypted and thus insecure. The security can further be increased by sending the data in several packets – if possible via different channels.

The data are decrypted by an inverse function E^{-1} which compiles the encoded message M_K into the plain text M with the same key K[43].

$$E^{-1}(K, M_K) \rightarrow M$$

or:

$$E^{-1}(K, E(K, M)) = M$$

Because of the difficult exchange of keys this method does no longer meet the requirements of an electronic communication.

Asymmetrical encoding methods or methods with public keys use different keys for the encryption and decryption. These keys are the key K_e on the sender side and key K_d known on the receiving side [96]. The encryption function E differs from the decryption function D.

$$D(K_d, E(K_e, M)) = M$$

The key exchange problem does not occur since K_d cannot be derived from K_e – or only at great expense. A simple example is a method where an integral number is partitioned into its prime factors: While the product can easily be partitioned into its prime factors the reverse is possible but only at great expense[44]. This expense depends on the size of the number – i.e., on the length of the key. The recipient generates of pair of keys (K_e, K_d). The private or secret key K_d remains at the sender while the public key K_e is available to all potential senders. This explains the often used term *public key system*. The sender encrypts his message with the public key, sends it via the transmission channel to the recipient who, by means of the private key, converts the message into the plain text.

Practical methods are different with regards to their algorithms, the length and distribution of their keys. One example is the *RSA method* (Rivest, Shamir and Adelman), which is based on the above-mentioned prime number partition [99]. The length of the keys is independent of the algorithm and determines the security degree. An efficient computer needs several months to partition currently used keys with a length of 128 bit into their prime

[43] The functions E and E^{-1} *may* be identical, e.g., the bit-wise XOR link of the message with the key.

[44] Each natural number n can be represented as potence of prime numbers P: $n = \prod_{i=1}^{k} P_i^{e_i}$; e.g., $1024 = 2^{10}$, $7007 = 7^2 \times 11^1 \times 13^1$
In case of large numbers the partition is very complex.

numbers. On the opposite, the product of the prime factors can be calculated by some multiplications. Principally, K_E can be calculated out of K_D. The required effort – above all the required time – increases exponentially to the length of the key so that keys of an adequate length are always secure enough. Public keys may be sent directly from the recipient to the sender via insecure channels or be published via central services. With regard to security the asymmetrical methods with their open systems prevail.

7.8.2 Integrity assurance

On its way from sender to recipient a message must not be forged by a third party. This *integrity assurance* protects the data against external affects. Order quantities, for example, which initiate a retrieval in a WMS, could be altered during their transmission. As a consequence, the complete inventory could be delivered and would not be available for other customers until the situation would be cleared and the stocks are returned.

The integrity can be assured by encryption. In a system with symmetrical keys the hacker has to have two secret keys, one to read and one to change the message. In asymmetrical systems the hacker has to have the recipient's secret as well as his public key.

7.8.3 Authentication

The integrity of a message can be assured by encryption. In asymmetric system everyone can sent a message to its owner and prevent to be someone else if he/she is in possession of the public key. For example, he/she can easily send orders to a supplier under the name of another customer. The supplier, however, has to be sure that the sender of the order is identical to the sender stated in the order. This *authentication* problem can also be solved by means of asymmetrical keys. There is quite a number of *authentication protocols* [77] which either call for a cental. reliable instance or the mutual knowledge of the public key.

The second case is described in a simple example with the participants A and B. A wants to communicate with B and agrees upon a *session key* K_S valid for the duration of the session. This session key has to be exchanged via a secure channel. For this purpose, an asymmetrical pair of keys is defined: (K_{EA}, K_{DA}) for A and (K_{EB}, K_{DB}) for B. After the authentication this session key is known only to both participants who now use it for a symmetrical procedure. An advantage of symmetrical systems are their considerably faster algorithms.

1. A randomly chooses a number R_A which he/she encrypts with the public key K_{EB} of B. A sends the result$E(K_{EB}, R_A)$ to B.
2. B decrypts the message with his/her secret key K_{DA}.

3. Since only the owner of this key can decode the message the random number chosen by A is returned to A as a "proof". In addition to this, B also generates a random number and a session key. B encrypts the received random number R_A, his own random number R_B and the session key K_S with A's public key K_{EA}. B sends the result $E(K_{Ea}, (R_A, R_B, K_S))$ to A.
4. A decodes the message with his/her secret key. When the received random number R_A corresponds with the number sent in step 1 B is reliable and the procedure can go on. If not, the procedure is terminated.
5. A operates with the session key $E(K_S, R_B)$ which he received from B. He/she encodes the random number generated by B with the session key and sends the result $E(K_S, R_B)$ to B.
6. B decodes the received message with the session key. When the decoded random number R_B corresponds to the number generated in step 3 A is reliable and starts to communicate.

For this method each communication partner has to know the public key of the other and the keys have to be assigned *reliably* to a participant. Public keys are managed in so-called *trust centers* which guarantee for the reliability of each participant.

An alternative to this are logical networks which set up decentrally a reliability variable. When A trusts the communication partner B and B trusts C then A also trusts C although to a somewhat lesser degree. Thus, the trustworthiness among the participants of a communication network can be propagated dynamically.

7.8.4 Authentication and electronic signature

It may be important for the sender as well as the recipient of a message to proof his/her authenticity. Here, we describe the basic principle of just *one* method based on asymmetrical keys. A prerequisite for this method is that in addition to the above characteristic $D(K_d, E(K_e, M)) = M$ this method can also be performed with "swapped" keys:

$$E(K_e, D(K_d, M)) = M$$

Thus, messages encoded with the private key can be decoded with the public key of the sender. For this purpose, the sender at first encodes the message with his/her secret key and then with the public key of the recipient. In the first step the message can only be decoded by the recipient with his/her private key. In the second step the recipient decodes the message with the public key of the sender. Since this key is only known to the sender, the recipient can be sure that the message was actually generated by him/her and that he/she cannot deny its existence and contents. The recipient cannot change the contents of the message afterwards because in case of a conflict

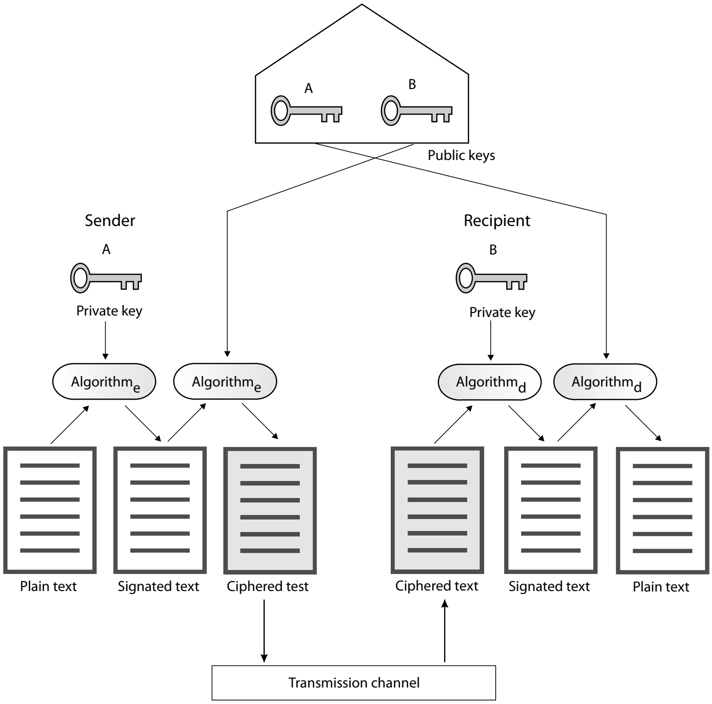


Figure 7.33. Basic signing and encoding of a document

the sender can request the message encoded with his/her private key as a proof. This cannot be manipulated by the recipient without the key. The basic procedure is explained in figure 7.33.

Since the expenditure for this two-step encoding is very high faster methods are used in practice which are also based on this principle and require the mutual knowledge of the public keys and their reliability.

8. Realization of Warehouse Management Systems

It takes about nine months to implement a WMS[1]. This shows that the introduction of a WMS cannot be compared to the installation of a software application like word processing, but has to be seen as a project[2]. Furthermore, the smooth operation of the warehouse during this time cannot always be guaranteed. Malfunctions may occur for different reasons (e.g., difficulties with the adaptation of new work processes, the operation of new software or the integration of new hardware). These malfunctions cost a lot of time and money and have to be minimized as far as possible.

A careful planning prior to the project start and a targeted implementation are, thus, indispensable. This chapter describes the systematic implementation of a WMS which basically corresponds to the introduction of a software of the same in size industry and trade (cf. Table 8.1).

In the following the single steps of the project are described including the respective requirements, activities and results. Two project types can be differentiated according to the initial situation:

1. The replacement of an existing WMS.
2. The introduction of a new WMS in line with the construction of a new warehouse.

In both cases, the project is carried out in the same way; however, additional restrictions have to be considered when an existing warehouse is replaced. For this reason, the following sections focus on such a situation. Should the procedure differ considerably for the new construction of a warehouse this will be mentioned explicitly.

The basic procedure is defined in the German standard VDI 2523 (Project management for logistic systems in material flow and warehousing).

[1] Source: International Market Survey Marktstudie WMS, http://www.warehouse-logistics.com

[2] A project which requires special organizational measurements. Generally, a project is limited in time, has a defined goal, is carried out once and is cross-sectional.

Table 8.1. Course of a project

Project step	Working point
Requirement definition	Actual analysis Weak point analysis Development of a target concept
Preparation of tender documents	Definition perform. characteristics Preparation perform. specs Completion of tender documents
Order placement	Preselection supplier and qualified shipment Of tender documents Comparison of offers Presentation of offers Selection of supplier
Realization	Preparation of performance specifications Realization Project management during realization / Quality assurance
Commissioning	Laboratory phase Change from old to new WMS Training
Acceptance	Performance test / function test / Availability check Simulation of malfunctions / test of emergency strategies Formal acceptance

8.1 Requirement definition

The project phase *requirement definition* is marked above all by the realization that the existing WMS no longer meets the requirements and/or the status quo has to be improved.

There are various reasons for the introduction or revision of a warehouse management system whereas an increase in efficiency is the major one. Other reasons are, for example:

– Introduction of a new warehouse technology (e.g., automatic high-bay warehouse) or of a new article to be stored (e.g., hazardous goods)
– Rationalization or increase of the key performance indicators by introducing new technologies (e.g., increase of the manual order-picking performance by pick-to-light)
– Use of the warehouse for different owners of the goods (multi-client capability)

- Integration of additional value-adding warehouse activities (Value Added Logistics, VAL)
- Outsourcing of logistic services
- Changed warehouse processes (e.g., improved quality assurance and documentation)
- Changed order structure or ordering behavior (e.g., because of a continuously increasing picking performance because of smaller quantities per order which is not efficiently covered by the system)
- Integration of e-commerce solutions (e.g., integration of an online shop)
- Changed requirements on the performance of the WMS (e.g., improved ergonomics)
- Securing the future by replacing outdated hardware and software (e.g., introduction of new databases)
- New interfaces to superordinate ERP and PPS systems (e.g., introduction of an ERP system on a company-wide level)
- etc.

8.1.1 As-is analysis

Because of the many requirements put on a new WMS, the coordinated and exact documentation of the status quo is of special importance. It is based on the as-is analysis and consists of the following steps:

Recording of the business processes and the information flow The aim of a business process record is to completely document all business processes[3] in a warehouse, mostly in the form of a business process chain. In addition to the physical activities also the information flow and the use of documents are recorded so that, for example, media breaks (change between two different media , e.g., EDP and paper) which affect the performance can be identified.

Recording of the existing performance of the WMS To get a clear impression of the performance of a WMS all functions and the related business processes and procedures are analyzed as well as the main response times (mask structure, database access, batch calculation, etc.).

Documentation of the interfaces to superordinate (ERP, MIS, etc.) and subordinate (WCS, MFC, etc.) systems When the interfaces between WMS and superordinate and subordinate systems is documented the complete data flow of the WMS is recorded. This documentation includes the type and number of external systems, the used protocols as well as the

[3] A business process is a series of activities covering different functions which creates a value for the customers and the company.

transmitted input and output data, the storage location for master data and the hierarchical design of data and information flows.

Description of the structure of relevant data The description of the structure of relevant data (e.g., article number, order number, batch, etc.) gives a complete overview over the structure of all important data, objects and number ranges like article master data, serial number, number of the loading aid, etc. (cf. Table 2.13, p. 57).

The aim of the as-is analysis is the continuous recording and description of all warehouse-related processes and systems.

8.1.2 Weak-point analysis

During a weak-point analysis all data of the as-is analysis regarding potentials for improvement are studied while the weak points are divided according to the work steps defined for the as-is analysis. Weak points are, for example:

Business processes and information flow in the warehouse By means of the process chain analysis[4] (PCA) weak points can be identified within a business process (e.g., inconsistent processes for nearly identical tasks or no defined processes for exceptional cases) and the information flow (e.g., media breaks).

Scope of performance of the WMS The WMS does not have the capacity required to ensure the smooth operation (e.g., no support with regard to the order sequence (batch optimization) by the WMS).

Interfaces to subordinate and superordinate systems Incomplete descriptions of interfaces recorded during the as-is analysis (e.g., incomplete description of the ERP interface).

Structure of relevant data The structure of the existing number ranges corresponds no longer to the actual requirements or needed number ranges are not available in the WMS (e.g., a serial number for tracking, lack of a corresponding data field in the WMS).

The aim of the weak-point analysis is to identify all weak points and to derive potentials for improvement.

[4] A process chain can be defined as a coordinated, i.e., process-oriented, sequence of single sub-processes (process chain elements). At first, the relevant process chains are recorded and analyzed with regard to their potentials (e.g., reduction of time and costs). ([45]; [46], p.162ff)

8.1.3 Development of a target concept

Before the target concept is developed it has to be decided whether the existing WMS should be replaced or not. In addition to the improvement potentials this decision should also be based on following aspects:

– Economic situation of the company
– Strategic orientation of the company
– Prognoses (e.g., macroeconomic, branch-specific, etc.)
– Feasibility study and risk analysis[5]

If it is decided to replace the existing WMS by a new one a target concept is developed. This concept is based on the results of the as-is analysis and the weak-point analysis and defines the requirements on a new WMS under consideration of the improvement potentials. According to the classification set up by the weak-point analysis, the requirements are as follows:

Business processes and information flow in the warehouse All business processes concerning the warehouse are clearly defined. The information flow and the use of documents are documented completely.

Scope of performance of the WMS The capacities of the WMS (functions and related business processes and procedures) complete meet the actual and planned requirements.

Interfaces to subordinate and superordinate systems The complete data exchange of the WMS is documented as well as all included external systems, the used protocols and the transmitted data. Furthermore, it is clearly defined which system has which date as master date.

Structure of the relevant data All required data, objects and number ranges are completely defined (cf. Table 2.13 (p. 57) and Table 8.2).

Finally, a set-up schedule is prepared for the target concept which shows the rough milestones and the required time (cf. Table 8.3).

8.2 Preparation of the tender documents

During the project phase is performance specifications, technical specification and call for tenders the elaborated target concept is formally written down under consideration of valid standards, guidelines and laws as well as company-specific regulations. The goal of this project phase is the preparation of tender documents.

[5] The aim of the risk analysis is to identify and evaluate as early as possible potential internal and external project risks.

Table 8.2. Target concept: Example of the structure of a storage bin file

No	Field	Description
1.	Warehouse	Identifies the warehouse
2.	Area	Identifies the bin location
3.	Aisle	Identifies the aisle
4.	Side	Defines the side of the aisle
5.	X	X-coordinates of the shelf
6.	Y	Y-coordinates of the shelf
7.	Z	Z-coordinates of the shelf
8.	MaxHeight	Maximum usable height of the storage bin
9.	IDLocProp	Storage bin ID
10.	IDTSUTypeGroup	Group of TSU which can be stored in this bin
11.	IDTSUType	TSU type currently stored in this bin
12.	Quantity	Number of TSU stored in this bin
13.	ResQuantity	Reserved quantity – Number TSU reserved for retrieval
14.	RepQuantity	Replenished quantity – Number TSU on the way as replenishment for this bin
15.	IDBlock	Blocking ID
16.	InvTS	Date of the last inventory
17.	InvIDOperator	Operator who carried out the last inventory
18.	ZCrossingTS	ZeroCrossingTimeStamp – Time of zero-crossing

8.2.1 Definition of the key performance indicators

In the performance catalogue the key performance indicators (KPI) are listed, for example:

– Orders, items, grips per time interval (day / week / month)
– Item, grips per article / assortment / client
– Orders, volumes, weight per customer / type of shipment
– etc.

To avoid an under-dimensioning of the WMS the KPIs have also to consider long-term developments (e.g., extension of the assortment or introduction of performance-related wages for pickers) in addition to current requirements. The KPI define the requirements on a new WMS and are an important part of the tender documents.

Table 8.3. Target concept: Example of a set-up schedule

Month Project step	1	2	3	4	5	6	7	8	9	10	11	12
Process analysis	•											
Requirement definition	•	•										
Requirement specs		•	•									
Preselection supplier				•								
Call for offer				•	•							
Selection supplier					•							
Project definition					•							
Functional requirements					•	•						
Realization						•	•	•	•	•		
Installation / Test									•	•		
Set-up										•	•	•
Acceptance												•

8.2.2 Preparation of the technical specifications

Another important part of the tender documents are the technical specifications:

> "The technical specifications of a project summarize the requirements on the system to be developed, in this case the program/program system. This requires:
> – a feasibility study and
> – an extensive analysis of the risks which occur or may occur during the development" ([63], p. L-1).

The main aspects for the preparation of a WMS requirement specification are ([6], p. 63):

Task The requirement specification describes in detail all technical requirements which have to be met by the WMS from the customer's point of view.

Addressees The usual recipients of the technical specifications are the marketing experts of the WMS provider and the potential project manager as well as the application specialists.

Content The problem to be solved, not the solution, is described on sufficiently high abstraction level (WHAT and HOW).

Form A clear numbering of the single requirements makes it easier, for example, to compare the offers and prepare the functional specifications

Time The technical specifications are the first formal document describing the main requirements on the WMS.

Scope Since the specifications focus on the fundamental requirements they can be limited to a few pages.

The above-mentioned aspects result in the following rough scheme:

A) Basics (Short, unfunctional description of the project)
  - Task
    General strategic description of the task (e.g., replacement of the existing WMS by a new one).
  - Reasons
    Reasons and necessities of the tasks (e.g., insufficient capacity of the existing WMS to meet all warehouse requirements).
  - Project target
    General description of the project target (e.g., functional support of all warehouse activities by a new WMS whereas the use must be able to parameterize this system to future events.). The planned key requirement parameters are also listed.
  - Project environment
    Short description of the environment (e.g., headquarters with production and intermediate warehouse of the company in city A, distribution centers in the cities B and C).

B) Scope of functions (description of all functions)
  - Limitation to ERP and MFC and other software
  - Functions
    List and description of all indispensable functions and services (e.g., batch capability).
  - Optional functions
    List and description of all functions and services (e.g., online language switch without renewed login) which would help to operate the warehouse but are not absolutely necessary (e.g., because of increasing costs).
  - Special functions
    List and description of all functions and services which are not related to the daily warehouse operation but are necessary for its administration (e.g., compatibility/upgradability: The user can represent other warehouses in the WMS without support from the WMS provider).
    Another special function other than the representation of uncommon functions is the description of data to be transferred from the WMS

The content of this exemplarily requirements specifications for the choice of a WMS is represented in the form of a directory (cf.Fig. 8.1). The following standards and guidelines should be considered while preparing a requirement specification:

Contents WMS requirements specification

1. Task
   1.1 Reason
   1.2 Project goal
   1.3 Project environment
   1.4 Project key data
   1.5 Standards and guidelines
2. Process analysis
   2.1 Structural organisation
   2.2 Data presentation and quantity structure
      2.2.1 Article-related master data
      2.2.2 Article data storage and transport system
      2.2.3 Goods access
         2.2.3.1 GR zones
         2.3.3.2 Storage strategy according to storage area
         2.3.3.3 Incoming goods inspection
      2.2.4 Order-picking
         2.2.4.1 Number of released orders
         2.2.4.2 Picking method
         2.2.4.3 Odd lot check on retrieval position
         2.2.4.4 Retrieval check
         2.2.4.5 Replenishment strategy
      2.2.5 Goods issue
         2.2.5.1 Goods issue zones
         2.2.5.2 Shipping methods
         2.2.5.3 Loading ch
   2.3 Goods issue process
      2.3.1 Overview standard procedure
      2.3.2 Order data
      2.3.3 Handling of goods inputs
         2.3.3.1 Deliveries
         2.3.3.2 Determination of storage location
      2.3.4 Storage transport
      2.3.5 Incoming goods booking
   2.4 Order processing procedure
      2.4.1 Order data
      2.4.2 Order-picking
         2.4.2.1 Release of orders for picking
         2.4.2.2 User guidance order-picking
         2.4.2.3 Generation of replenishment orders
      2.4.3 Extraordinary business events
         2.4.3.1 Cancellation of picking order
         2.4.3.2 Error in picking order
         2.4.3.3 Requirements on QA
      2.4.4 Goods output
   2.5 Information and reporting
      2.5.1 ARticle stocks
      2.5.2 Warehouse data
      2.5.3 Order data
      2.5.4 GR data
      2.5.5 Movement data, movement statistics
3. Definition of requirements
   3.1 Target criteria
   3.2 Desired criteria
   3.3 Targets
4. Scope of delivery and services
   4.1 General offer and delivery terms
   4.2 Hardware
   4.3 Software
   4.4 Documentation
   4.5 Other deliveries and services
      4.5.1 Training and instruction
      4.5.2 Warranty, maintenance
      4.5.3 Performance specifications, availability specifications
   4.6 Project execution
   4.7 Proof of performance and availability
5. Costs
6. Secrecy agreement
7. Copyright
8. Images
9. Abbreviations

Figure 8.1. Example of the directory of a requirement specification

- VDI 2519 (Procedure for the preparation of technical specifications)
- VDI/VDE 3694 (technical specifications for automation systems)
- VDI 3969 (WMS interfaces to superordinate systems)
- DIN ISO 9126 (Software quality characteristics)

In addition to the above-mentioned standards and guidelines the following standards should be considered when defining the availability:

- VDI 4004 Sheet 4 (Reliability parameter - availability parameter)
- VDI 3649 (Use of the availability calculation for conveyors and warehouse facilities)
- VDI 3581 Draft (Availability of transport and warehouse facilities as well as of their sub-systems and elements)

As a conclusion it can be said that the technical specifications should be structured sensibly. Offers should refer directly to the technical specifications.

8.2.3 Completion of the tender documents

The tender documents are completed among others by the following points:

- List of the general purchasing conditions
- Addition of a secrecy agreement
- Information about possible penalties (also: right to amendment)
- Provision of a binding template for the structure of the offer
- A binding template for the price list (cf. Table 8.4).

Table 8.4. Example of a price list

Services	Costs in EURO
Basic module	
Main function 1-n	XX.XXX,XX
Function 1-m	
Option 1-n	XX.XXX,XX
Function 1-m	
Hardware	
Device 1-n	XX.XXX,XX
Component 1-m	
Training	
Subject 1-n	XX.XXX,XX
Maintenance	
Maintenance agreement 1-n	XX.XXX,XX
Services 1-m	

8.3 The placement of an order

Before an order is placed the providers have to be preselected, the tender documents have to be sent and, if required, the offers have to be presented.

8.3.1 Preselection of providers

To prepare for the *preselection of providers* and to compare the offers the technical specifications have to divided into the following criteria:

– Criteria which are indispensable for the smooth operation of the warehouse, e.g., multi-client capability for logistic service providers, become exclusion criteria: If this requirement is not met by the product of the WMS provider the product is immediately excluded from the selection process.
– Criteria which support the operation but are not indispensable become so-called optional criteria: This function supports the operation but if it is not offered this WMS is not automatically excluded.

The list of offers to be compared can be reduced drastically by addressing only providers who meet the minimum requirements (the so-called qualified sending of tender documents). An example of a checklist for these minimum requirements is shown in Table 8.5.

The preselection of WMS providers is supported by a database mentioned in [74]. Table 8.6 shows the four-step process with the aim to generate a list of providers which meet the requirements as basis for the qualified mailing of the tender documents.

8.3.2 Comparison of offers

Since WMS providers do not always offer the performance of the functions described in the technical specifications, but those covered by the WMS, the incoming offers have to be compared according to the checklist shown in Table 8.7.

Each offer is tested systematically, which in most cases leads to questions which have to be cleared. For this purpose, it may be necessary to extend or detail the offer at hand. Table 8.8 and 8.9 show examples of the structure of a comparison of offers.

8.3.3 Offer presentation

Generally, providers or offers cannot directly be evaluated by a direct numerical comparison, and this method would not be adequate for the choice of a WMS. A meeting of customer and consignee provides further helpful information, e.g., in the form of an offer presentation.

Table 8.5. Checklist minimum requirements

Which kind of warehouse projects should the provider have carried out? E.g., • Automatic mini-load warehouse • manual container / small parts warehouse	
For which kind of special goods should the provider have carried out projects? E.g., • Piece goods • Coil • Long goods • Bulk goods	
For which kind of warehouses should the provider have carried out projects? E.g., • Order-picking warehouse • Bonded warehouse • Cold store • Hazardous goods warehouse • Multiple warehouse: central warehouse with regional stores	
Is a software required which is suitable for shippers?	
Is a software required which is suitable for parcel services?	
Is a software required which is suitable for logistic service providers?	
Is a batch management required?	
Is a serial number management required?	
Should the system by multi-client capable?	
Should the system be a Mehrlager?	
Should the dock / yard management be supported by the WMS?	
...	

This gives the WMS provider the opportunity to personally present his services as well as the excellence of his product and his company. Furthermore, the procedure/methodology or the project schedule for the introduction of a WMS can be discussed, and the customer has the possibility to clear unanswered questions. The results and findings of this offer presentation are recorded in the form of a protocol.

On the customer's side the responsible decision-makers, the operative warehouse staff and the project manager (of the customer) should attend this presentation while on the provider's side at least the responsible marketing staff the potential project manager (of the consignee) should be present.

A standardized procedure makes it easier to compare and evaluate the presentations of all WMS providers, i.e.:

Table 8.6. Internet-based preselection of providers

Selection of WMS[a]		Contents
1.	Definition of exclusion criteria	The user gets from the system a list of possible exclusion criteria. These are determined according to the requirement profile.
2.	Shortlist	The system generates a list of all products which meet the defined exclusion criteria.
3.	Weighting	The complete data stock is divided into function groups which can be weighted individually by means of the rough weighting. Additionally, all questions and answers can be weighted explicitly up to 1000 single aspects in the scope of the detailed weighting. With this detailed weighting the user can exactly specify his/her requirements.
4.	Result	As a result the user gets a list of all eligible products with the degree of performance of the superordinate function groups. Furthermore, the actual answers of the providers concerning each function group can be viewed.

[a] Source: International market survey WMS, http://www.warehouse-logistics.com

A) Each participating member of the customer's staff acquaints himself with the functions stated in the offer; this makes it easier to complete the evaluation sheet (Table 8.10) during the presentation.
B) Each participant makes a short and precise remark (e.g.: Requirements are met; requirements are not understood; requirements not mentioned; unacceptable, etc.) about his personal impression of the functions in the column "Status offer" of the evaluation sheet Table 8.10

It is recommended to complete the evaluation sheet (Table 8.10) as follows:

A) There is one evaluation sheet for each bidder.
B) In the column "Status offer" the evaluations of the functions — based on the offer — are listed which have been made during the preparation of the presentation.
C) During the offer presentation each participant enters his personal impression into the column "Comment".
D) A short and precise remark about the personal impression during the offer presentation in the column "Status presentation".
E) Finally, each function is evaluated according to the school mark principle in the column "Value".

The offer presentation has the aim to clear all unanswered questions of the customer as well as the WMS provider.

Table 8.7. Checklist comparison of offers

Offer of basic modules (Are all required basic modules offered?)	
Offer of options (To what extent are the functions offered?)	
Standardization and release capability of the offered software (Are the basic modules and options part of the standard[a] and are there adequate references or have they to be implemented individually?)	
Licence models for server and client software	
Offer of training?	
Offer maintenance and care	
Detailed description of the expenditure for the individual programming	
Requirements of the WMS provider on the customer (e.g., How many employees have to be allocated for this project, which precautions have to be made?)	
Daily or hourly rates and travelling expense: Differentiated according to qualification and activity of the employee	
Scope of the offer (Is more offered than required? Why is more offered? Is this "more" necessary from the customer's point of view?)	
Check of the general terms and conditions of the provider	

[a] If the function is part of the standard, it certainly has been implemented several times with success. A quick and faultless implementation is more likely than in case of an application-specific solution.

8.3.4 Selection of a provider

All findings gathered during the presentations are used to evaluate the offer. After the final evaluation of all offers, the providers are listed according to their priority. Intensive negotiations can be started with bidder 1. If these negotiations fail (e.g., for cost reasons) negotiations with bidder 2 follow, etc. In the end that WMS provider has to be chosen who will provide and implement the new WMS system.

8.4 Implementation

During the implementation the requirement specifications are prepared and the system is implemented supported by a quality assurance.

8.4.1 Preparation of the technical specifications

The requirement specifications describe *how* the technical specifications are fulfilled. It sets the basis for all activities during the implementation and should be an integrated part of the agreement between customer and consignee.

Table 8.8. Evaluation of offers

Function / Services		Seller A	B	...
Data presentation and quantity frame		⊗	⊗	
Goods-related master data		⊗	⊗	
Article master data		+	+	
Packaging unit per article		+	(+) 1	
Table with types of goods		+	- 2	
Master data warehouse and transport systems		⊗	⊗	
Master data warehouse	KO	⊗	+	
Master data fork list truck		⊗	- 3	
Procedure goods receipt		⊗	⊗	
Goods receipt		⊗	+	
Goods delivery		+	(+) 9	
Determination of storage bin		⊗	+	
Storage transport		(+) 28	+	
Goods receipt posting		⊗	⊗	
Accounting WMS		(+) 29	(+) 10	
Input from production		(+) 31	+	
Procedure customer order execution		⊗	⊗	
Order-picking		⊗	⊗	
Order release for picking		⊗	+	
Manually		+	⊗	
Automatic without restrictions		(+) 32	⊗	
Automatic with restrictions		(+) 32	⊗	
Operator guidance order-picking		- 34	+	
Generation of supply orders		- 34	+	
Goods issue		⊗	+	
Information and reporting		+	+	
Data transfer OLD-WM	KO	-	-	
2-phase order-picking	KO	-	-	
Prioritization	KO	-	-	
Compatibility RF		(+) 2, 6	-	
Maintenance agreement		-	-	
Training		+	+	
requirement specifications		-	-	
Availability specifications		-	-	
Project management		-	-	
...				

Legend

KO	Exclusion criterion; an empty field marks an optional criterion.
+ / -	Requirement is stated in the offer and met / not met.
-	Requirement is stated in the offer but not met.
⊗	Requirement is not stated in the offer.
32	Topic in Table 8.9.

Table 8.9. Comments on the offers

Topic	Offer page	Technical specifications chapter	Comments company A
2	5	2.3	NT-computer and console software stated as additional costs.
6	8	2.5	Why is it recommended to replace older devices? Are there problems with the use of older devices or is this a general opinion?
28	18	2.3.4	"Consolidated transports" is an exclusion criterion. Stated as Add-On "combined transports"; no exact price given.
34	19	2.4.3.2-3	"Operator guidance order-picking" and "generation of supply orders": Both have to be adjusted. Already included in the quoted price?
...			

Table 8.10. Evaluation sheet

No.	Functions	Status offer	Comment	Status presentation	Value
1	Equipment master data				
2	Storage strategies				
3	Supply order-picking				
...					

> "The requirement specifications mainly describe how the tasks specified in the technical specifications have to be solved. It describes the performance of the system to be developed and specifies how and by which means the tasks have to be solved. Before this background the requirement specifications sets the basis for the acceptance....Customer and consignee shall prepare a performance specification based on the technical specifications as basis for the agreement..."([63], p. L-1).

When the provider and the customer communicate continuously during the preparation of the requirement specifications unclear terms can be defined and summarized in a glossary. Thus, different interpretations of the specified performances can be minimized what allows for a more exact implementation.

> "The requirement specifications are a refinement of the technical specifications.... Based on the requirement specifications and the glos-

sary a formal product model is developed which has to be specify the requirements completely, consistently and clearly" ([6], p. 100).

The aspects of the requirement specifications principally equal that of the technical specifications (cf. section 8.2.2) plus the complete, consistent and unambiguous description of the fulfilment of the requirements defined in the technical specifications. Should new unpredictable findings occur during the implementation, these should be integrated into the requirement specifications with the agreement of the customer responsible consignee.

The consolidation of the following standards and guidelines facilitates the preparation of the requirement specifications and their observance may become part of the contract, if necessary:

- VDI 2519 (Procedure during the preparation of the requirement specifications)
- VDI/VDE 3694 (requirement specifications for use in automation systems)
- ANSI/IEEE 830-1998 (SRS Software Requirements Specifications)

The following rough scheme of a performance specification includes the main requirements:

- Basics (short, unfunctional project description)
  - Tasks
    Basic (business) description of the tasks, e.g., replacement of the existing WMS by a new WMS.
  - Reasons
    Reasons and necessities leading to the task, e.g., new activities in the warehouse can no longer be performed by the old system
  - Project goals
    General description of the project goal, e.g., functional support of all warehouse activities by a new WMS whereas this WMS has to be adjustable by the user to future events. The performance parameter have also to be stated.
  - Project environment
    Short description of the environment, e.g., headquarters with production and intermediate warehouse of the company in city A, distribution centers in the cities B and C.
  - Time schedule
    Mention of all milestones with key dates.
- Scope of functions (description of all functions)
  - Delimination to ERP and MFC and other software: Positioning of the WMS within the system landscape.
  - Functions: List and description of all functions and services.
  - Interfaces: Description of the necessary interfaces and protocols.
  - Data transfer: Type and method of the data transfer with information about the data to be transferred.

– Training (date, place and volume as well as target group of each single training)
– Start-up (date and procedure of the start-up)
– Acceptance (form and scope of the acceptance with precisely defined performance key values and a list of performances to be provided by the customer and the consignee)

It is the aim of this project phase to exactly and unambiguously define all requirements (services, activities and functions) including their scheduling which are necessary to achieve the project goal. The requirement specifications are approved by the customer and it is binding for both parties as part of the contract.

8.4.2 Realization

During the realization the tasks defined in the requirement and technical specifications are fulfilled to complete the product. It depends on the kind of WMS:

– Implementation (Installation of the software on the provided hardware)
– Customizing (Adaptation to the existing interfaces and activation of the necessary program modules)
– Parameterization (Adaptation of the software to the customer requirements by means of parameter and switches)
– Individual programming (Programming of individual functions)

All changes of the technical and requirement specifications which occur during the realization have to be documented and communicated between customer and consignee. Possible reasons for these changes may be new findings (e.g., from simulation), missing topics in the technical and the requirement specifications or new requirements (e.g., additional statistical information requests). The documentation should include, among others:

– Which service/function drops out; is added; is altered?
– This leads to which cost-effective changes?
– This leads to which time effects (scheduling of project milestones)?
– Which priority has this change? Has it to be completed before the start-up? When the change can be performed later by when has it to be performed at the latest?
– Does the change have any effect on the acceptance?

A documentation of changes avoids disputes during the acceptance because the differences between the functions described in the technical and the requirement specifications and the implemented functions can clearly be demonstrated.

8.4.3 Project management / Quality assurance

During the realization the usual controlling measures should be taken in the scope of a supporting project management:

Supervision of the milestones Has the time schedule of the single milestones been observed, are there project backlogs or is the total lead time in danger because of delays on the critical path?

Resource control Are efficient resources available? Would an increase in resources significantly improve the productivity? Has the the planned budget kept?

Furthermore, the following quality assurance measures have to be considered as part of the quality management:

– Quality management (QM) includes

  "all activities of the management which define the quality policy, goals and responsibilities and realize it in the scope of the quality management system by planning, controlling, assuring and improving the quality" [23].

– The general term QA describes

  "all planned and systematic activities which are realized within the quality management system and are presented as required to create an adequate trust into the ability of the unit to fulfill the quality requirements" [23].

This step focuses on the fulfilment of the function defined in the requirement specifications within a given period of time with the planned resources and in the desired quality.

8.5 Start-up

The start-up consists of the test phase with the offline performance tests, the changeover from the existing to the new WMS and the parallel training courses. When the start-up has been finished successfully the new WMS is released for the *production process.*

8.5.1 Test phase

Prior to the test phase the old data stocks (e.g., article master data, warehouse movement protocols, etc.) have to be transferred. It may be necessary to develop, for example, special import programs for this purpose. The data transfer may be very time-consuming so that it should be planned in good time, especially when large data stocks have to be transferred.

During the test phase the new WMS is tested under laboratory conditions but with real data: The orders submitted by the WMS are not really executed by the picker but in a simulation (e.g., a test program directly enters the picked quantity into the database). On the one hand, the used data should reflect the current requirements on the other hand the client and server software should prove in offline tests that the system is able to cope with the peak loads of the last months as well as projected, future loads. If a material flow is simulated the software and simulation data have to be consistent.

8.5.2 Changeover from old to new WMS

The changeover from the old to the new WMS can be made in two ways:

Parallel operation During the parallel operation or *parallel run* [6] both system run simultaneously whereas at first only the old WMS is active. Then, the new system is activated and tested section by section and successively replaces the old system. This method is very safe because in case of failures the old system can be used. A disadvantage, however, is the high cost for the operation of two systems and the general problems which arise during the parallel operation of two systems.

Direct changeover In case of a direct changeover the old WMS is directly replace by a new one. The old system is switched off and the new one immediately starts to work. Usually, the direct changeover is chosen when the transition period is limited for business reasons, e.g., during the company holidays or outside the seasonal business. An advantage of the direct changeover are the low costs. However, it may come to considerable start-up problems or data inconsistencies.

8.5.3 Training

The early and careful training of the staff is of great importance so that the participating employees have to be set free. An ill-prepared training, above all after work hours or on weekends, may have a negative effect on the readiness of the staff to accept the new system. At the end of this project phase trained members of the customer's staff are able to manage and control the warehouse supported by the new WMS under consideration of the functions described in the technical and requirement specifications and to eliminate all failures which occurred during the start-up.

8.6 Acceptance

During the acceptance the functions and response times described in the technical and requirement specifications are controlled, failures are simulated and emergency strategies are checked. Acceptance is based on the German guidelines VDI 39777E and 3979 and concluded by the acceptance certificate.

> "**Acceptance** Legally defined procedure where the customer accepts the product. The accepted product becomes the property of the customer"([6], p. 1098).

8.6.1 Performance test

The performance test with the function test of the main parts and availability test comprises the uninterrupted operation (mostly over several days) under real conditions and full load. When the operation has to be stopped due to failures the test has to be repeated.

During the function test at first, it is checked if all functions described in the requirement specifications exist. Then, the proper function of each single element is tested individually. Special attention should be given to functions which deviate from the standard[6]:

– Does the function exist?
– Does the function meet the requirements?

Typical performance key values which are determined during the test are the duration of the mask buildup on the screen, the duration of a database access (e.g., display of master data and statistics) and the time for a database request (e.g., open picking orders) as well as the reporting time (e.g., article per client per month). Furthermore, the response times to the external/mobile terminals (e.g., stacker terminals) have to be recorded according to VDI guideline 3649.

8.6.2 Failure simulation and emergency strategies

A failure simulation simulates extraordinary situations as described in the requirement specifications:

– System behavior in case of an interrupted communication with the host computer
– Consequences of a power breakdown
– Strategy check (Which strategy applies when an article has to be retrieved but the bin is empty? How does the system react when an article is going to be stored in the wrong area, e.g., frozen goods in the hazardous goods area?)
– Instruction and documentation check (What has to be done in case of an emergency?)
– etc.

[6] Functions deviating from the standard are those functions which have been implemented individually for the project and have not be adapted to the project by parameters or customizing.

In failure simulation adequate emergency strategies which maintain the operation are of special importance. The tests should also include the restoration of data and the synchronization of the different sub-systems (e.g., WMS and ERP) after a breakdown.

All criteria, services and availabilities which are described in the requirement specifications and which are relevant for the acceptance are listed in the transfer protocol.

8.6.3 Formal acceptance

The single positions are accepted one after the other and signed by the customer and the consignee. Each fulfilment, each fault and each deviation is stated in the protocol. In case of faults and deviations it is discussed to which of the four fault categories they belong:

– The warehouse cannot be operated due to the fault or the functional deviance from the requirement specification (e.g., the safety light barrier or the control of the rack feeder do not work).
– There is only an expensive alternative to the faulty or deviating function (e.g., the scanner at the goods receipt does not work and each article has to be scanned manually).
– There is an acceptable alternative for a faulty or deviating function (e.g., some of the data are not recorded by the WMS during the automatic order acceptance by the host which has to be carried out twice a day. The operator has to check the order acceptance and arrange for a renewed input, if necessary).
– A fault or deviation does not limit the operation of the system (e.g., the print format of the supply note in the goods issue does not exactly meet the specifications).

They have the following effects on the acceptance and the decisions about the implementation of the WMS:

Category 1 No acceptance, no putting into operation.
Category 2 Clarification who has to bear possible additional costs incurred by the customer (e.g., additional staff has to be employed until elimination of the fault). It can only be decided from case to case if parts of the system are accepted with restraints and the WMS is put into operation.
Category 3 The system is accepted with restraints and can be put into operation. The unfilled points have to be settled with highest priority.
Category 4 The system is accepted with restraints and can be put into operation. The fulfilment of the unsettled points has to be scheduled by the customer and the consignee.

Minor deviations should not prevent the commissioning of the system. It has to be defined, however, up to when the fault or deviation has to be eliminated. If the number of faults of the categories 3 or 4 gets too high an acceptance should be reconsidered. The categories also imply the priority of the fault elimination.

The final formal acceptance of the total system according to § 12 VOB, part B and VDMA follows the faultless, unobjected performance test which has to be completed at least 3 months after the start-up. All shortcomings stated in the handover certificate have to be eliminated prior to the acceptance which has to be requested in writing by the consignee.

If there is no other legal agreement, the system is accepted solely by the written certificate of the customer; it is replaced neither by an earlier use, start-up or acceptance by an authority nor by the consignee's notice of completion.

The customer as well as the consignee have to deploy sufficient material and qualified staff for the acceptance and tests to guarantee for the smooth performance of the stated activities. In addition to this, they have to appoint qualified contact persons.

The acceptance generally constitutes the transfer of risks from the consignee to the customer and the start of the contractually agreed warranty period.

With the acceptance the project *implementation of warehouse management systems* is finished. All guaranteed features have been checked, the facilities meet the requirements, the acceptance certificate has no faults or gaps and has been signed by both contract parties. The WMS goes into service.

9. Structure of a WMS from the Example of myWMS

A large variety of different warehouse management systems are available on the market which differ with regard to

– their scope of functions,
– their interfaces,
– the required hardware,
– the required operating system,
– the operating concept

and to many other features as well as, last but not least, regarding the investment costs. Often, with his decision for a certain WMS, the warehouse owner binds himself for a long time to the provider and depends on him for the extension of functions and interfaces and the exchange of hardware.

A modular and open warehouse management system which is programmed in a widely accepted language independent of the hardware and operating systems improves the compatibility to third-party products and components and ensures the efficient adjustment to the rapidly changing market conditions.

Such a comprehensive WMS solution was developed at the Fraunhofer IML[1] with the name myWMS. This system is a *framework* based on the principles of a modular WMS. It provides a number of software modules and regulations for the software-related construction supporting a WMS.

In this chapter, a general data model is first described which can be found in many warehouse management systems and sets the basis for myWMS. This discussion is followed by the description of a classical WMS implementation and finally the elementary structures of myWMS are presented and demonstrated at a practical example.

9.1 Data model

The core of each WMS is a data model which sets the basis for the management of the different data stocks and data flows which occur during the operation of a warehouse. This data model represents the relevant practical parts in the form of software structures (*data structures*). The first modelling

[1] Fraunhofer-Institute for Material Flow and Logistics, Dortmund, Germany

step is the analysis of all WMS relevant data, their structures and formal description. Then, their interrelations have to be studied and specified.

9.1.1 Data container of the model

The numbered boxes in Fig. 9.1 represent the data containers of a WMS, in the following called containers. A container takes up logically consistent data with the same structure.

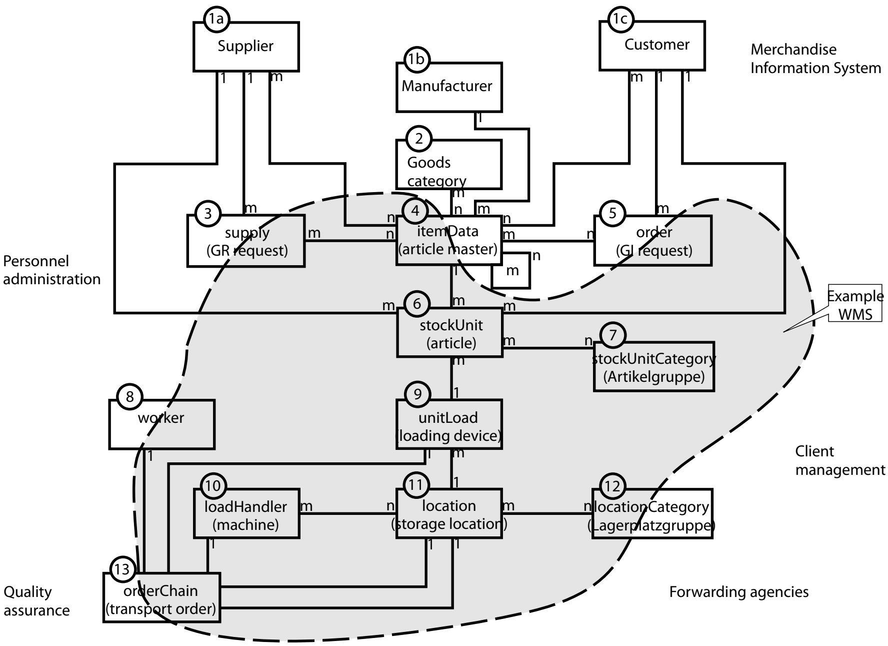


Figure 9.1. The data model of a warehouse management system with the adjacent systems. The *dotted line* separates the WMS from its environment. The entries in the data containers show the terms used in myWMS.

The exact data contents of a container does not necessarily have to be specified here so that, for example, the container 1c in Fig. 9.1, which represents the customer, may only hold a customer number or detailed customer data like surname, name, delivery address, billing address, telephone number and payment behavior. They just have to exist in the model; their meaning has to be defined and they have to represent reality.

- **Supplier:** The container 1a holds the supplier data
- **Manufacturer:** The container 1b holds data about the manufacturers of single articles

- **Customer:** The container 1c holds data about customers which are supplied via this warehouse
- **Article group:** Articles can be grouped within a warehouse according to different criteria. This is achieved by means of container 2.

The first three containers have been numbered with 1a, 1b and 1c. Since a supplier may also be a customer or a manufacturer a supplier (or vice versa), it would make sense to combine the containers 1a, 1b and 1c into one container which represents *a legal person* to avoid redundant data stocks. A WMS does not need detailed information about this person. Therefore, in Fig. 9.1, it is located outside the dotted line like the container for article groups.

Data containers which are relevant for a WMS and are required by external systems (e.g., personnel management, quality assurance and goods economy) (cf. section 9.1.3) are described in the following. The terms used in the diagram are generally used throughout the myWMS project (cf section 9.3).

- **Supply**: A supply is an outgoing order (and possibly its execution) of articles for the warehouse.
- **ItemData**: Container 4 holds the meta data for the articles. The *ItemData* describe the characteristics of the articles and not the pick-up unit in the warehouse which is associated with a serial number. They above all include data which are required for the correct storage and movement such as preference zone (A, B, or C), allowed light and temperature range, number of pick-up units per pallet or the hazardous goods category and weight (cf. section 2.4).
- **Order**: Container 5 holds incoming orders (customer orders) and is similar to the container *Supply* where just the customer and the supplier side are swapped.
- **Worker** This container holds data about the warehouse staff, e.g., of pickers. Above all information about the skills of workers and their authorization to operate certain machines - e.g., fork lift trucks - are important in a warehouse.
- **LocationCategory**: Locations can be grouped in container 12 where they can be allocated to a zone or storage group or categorized otherwise.
- **OrderChain**: Container 13 holds data about the single movements of the loading aids between the bin locations. It can also be seen as a kind of protocol.

The last container group is of relevance only within a WMS.

- **StockUnit**: Container 6 holds the data for those articles the meta data of which can be accessed via container 4 (*ItemData*). A device with a serial number substantiates its meta date. The TV set "look-out" with the serial number 405432100002, for example, is entered into the container **StockUnit** whereas in the container **ItemData** all data of the model

"Look-out" are described. Here, the concrete model only affects the stock. Similar approaches are also used in systems without serial numbers.
- **StockUnitCategory**: Here, concrete stock units can be built, for example by building batches or tour numbers of the supplier.
- **UnitLoad**: Container 9 describes the unit load which may hold stock units. A unit load may be loaded with mixed or single-type articles.
- **LoadHandler**: The load handler is used for the physical transport and for this purpose temporarily has to take up unit loads. Thus, a load handler is a "bin location" for a short time.
- **Location**: In container 11 the bin locations are recorded. A location may be occupied or free. In addition, a blocking mark can be set like for all other containers (cf. section 2.2).

The containers described up to now represent the typical warehouse data. Further containers can be added to this model, e.g., a container for the grouping of unit loads to classify unit loads with partially similar characteristics. This model is based on a warehouse with few types of unit loads, as it prevails in practice, and represents the differences in an attribute of the container UnitLoad. A real data model does not necessarily use all containers.

9.1.2 Data interrelations

While the containers in Fig. 9.1 describe the data of a WMS, the lines point out their interrelations. Several kinds of uni- or bilateral relations are possible and their multiplicity is marked by numbers or characters at the end of the connecting line.

- A 0 means that there is no relation in that direction. A 0 does not have to be shown to simplify the presentation. Relations with a 0 at both ends are called empty relations, hold no information and are not drawn.
- A 1 at one end of a connecting line shows that at this end exactly one element of the container is assigned to the container at the other side. For example, a telephone number is related to exactly one legal person (customer, manufacturer or supplier).
- An n at an end of a connecting line indicates that there are a maximum of up to n relations to this side. For example, a legal person may have several telephone numbers but there *may* also be *none*.

In case of a multiple relation in both directions, one side is marked with n and the other with m to indicate that the multiplicity may be different. For example, a relation $\underline{n \quad\quad m}$ between the manufacturer and the supplier container could be shown if it is relevant. This relation could show that a manufacturer uses the services of n suppliers, while a supplier sells the products of m manufacturers. A relation line $\underline{n \quad\quad m}$, thus, is the symbolic combination of a $\underline{n \quad\quad 0}$ and a $\underline{0 \quad\quad m}$ line.

In the following, all relation lines shown in Fig. 9.1 are listed and described in a table. The numbers of the respective containers are connected by one of the lines mentioned above and marked with a name.

- (1a $\overset{1\quad\;\; n}{\rule{2cm}{0.4pt}}$ 6) **Delivery:** This relation describes the delivery of articles whereas a supplier delivers several articles ($n \geq 0$), but one article exactly belongs to the delivery of one supplier. Reasonable relations cannot be differentiated from unreasonable ones with the data model alone. This relation, for example, allows empty deliveries but also deliveries without articles. The superordinate logic has to ensure that a supply contains at least one article.
- (1a $\overset{1\quad\;\; n}{\rule{2cm}{0.4pt}}$ 3) **Order (or notification):** This relation represents an order (or a notification). An order (a notification) is assigned to exactly one supplier while a supplier may be allocated to several orders (notifications)
- (1a $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Catalogue:** The catalogue of a supplier contains the meta data of different articles but an article may also be bought from different suppliers.
- (1b $\overset{1\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Item list:** An article which is represented by its meta data is manufactured by exactly one manufacturer who may, however, have several articles in his assortment.
- (1c $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Purchasing behavior:** The purchasing behavior of a customer can be described by a relation between item data and customer. A customer may buy different articles and an article may be bought by different customers.
- (1c $\overset{1\quad\;\; n}{\rule{2cm}{0.4pt}}$ 5) **Order:** A customer may submit different orders during the time but an order is always assigned to exactly one customer.
- (1c $\overset{1\quad\;\; n}{\rule{2cm}{0.4pt}}$ 6) **Serial number tracking:** With this relation the delivered article can be traced back to the customer and vice versa. A warranty claim, for example, can be settled by tracking the relation starting from **delivery** (to the supplier). This relation also closes the chain for a recall.
- (2 $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Goods classification:** The relation goods classification combines the meta data of articles with the same characteristics. This relation is of interest mainly in shop or merchandize management systems.
- (3 $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Storage order:** A delivery (or notification) initiates a storage order which is related to a quantity of articles. A delivery is not only related to the article itself but also to its item data.
- (4 $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 4) **Mixed-storage prohibition:** This relation is necessary because some articles must not be stored together (cf. Chapter 2).
- (4 $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 5) **Retrieval order:** Each customer order is assigned to a quantity of articles which is described by this relation to the respective meta data.
- (6 $\overset{m\quad\;\; n}{\rule{2cm}{0.4pt}}$ 7) **Article classification:** An article may belong to several categories and each category contains a variety of articles.

- $(\; 6 \; \overset{n \qquad 1}{\overline{\qquad\quad}} \; 9 \;)$ **Load:** With this relation the articles are assigned to the unit loads. An article can be on just one unit load[2], but this *can* hold several articles.
- $(\; 8 \; \overset{1 \qquad 0}{\overline{\qquad\quad}} \; 13 \;)$ **Work order,**
- $(\; 9 \; \overset{1 \qquad 0}{\overline{\qquad\quad}} \; 13 \;)$ **Transport object,**
- $(\; 10 \; \overset{1 \qquad 0}{\overline{\qquad\quad}} \; 13 \;)$ **Transport medium,**
- $(\; 11 \; \overset{1 \qquad 0}{\overline{\qquad\quad}} \; 13 \;)$ **Transport source,**
- $(\; 11 \; \overset{1 \qquad 0}{\overline{\qquad\quad}} \; 13 \;)$ **Transport destination:** These relations all start from *one* transport order and are unilateral. The relation *work order* assigns just one *worker* to the transport order, e.g., a stacker driver for a driving order or a picker for a picking order. In a fully automatic warehouse the relation *work order* describes the responsibilities. The relation *transport object* indicates which unit load is moved while the relations *transport medium* indicates the relation to the executing load handler. The relations *transport source* and *transport destination* are self-descriptive.
- $(\; 9 \; \overset{n \qquad 1}{\overline{\qquad\quad}} \; 11 \;)$ **Bin occupation:** Each unit load has to be assigned to exactly one bin. A storage bin may take up several unit loads.
- $(\; 10 \; \overset{m \qquad n}{\overline{\qquad\quad}} \; 11 \;)$ **Working area:** A certain bin location may be accessed by several machines, but one load handler may also serve several bin locations.
- $(\; 11 \; \overset{m \qquad n}{\overline{\qquad\quad}} \; 12 \;)$ **Zoning:** Different locations can be grouped in zones and each zone may consist of many bin locations.

Further relations can be generated, but are seldom required. The model may be extended, for example, by a $\overset{n \qquad m}{\overline{\qquad\quad}}$ relation between worker and load handler, which informs users about the skills and authorization to operate certain machines. The respective characteristics then have to be removed from the container *worker*.

The existing relations may also appear in another multiplicity, e.g., the relation *working area* as a $\overset{1 \qquad n}{\overline{\qquad\quad}}$ relation when the bin locations are accessed by exactly one load handler, e.g., a rack feeder. Similar restrictions may also be made concerning the *article classification* when the single category is built by the batch.

The relations may also be interpreted in another way. The relation *purchasing behavior*, for example, may occur as *purchasing prohibition*[3]. If *purchasing behavior* as well as *purchasing prohibition* are wanted as a relation between customer and item master both are shown by a separate line.

---

[2] Articles which are transported or stored without a unit load are assigned to a "virtual unit load", which is physically nonexistent.

[3] Such a purchasing prohibition for single articles and countries was stipulated, for example, in the CoCom–list (Coordinating Committee on East-West Trade Policy).

9.1.3 Interfaces

The containers which in Fig. 9.1 are crossed by the dotted line build the interface of the WMS. The data in these containers are relevant for the WMS as well as for the adjacent systems such as

- Personnel management
- Quality assurance,
- Forwarding agencies
- Client management and
- Merchandize management

They can be implemented in two different ways: A joint database helps to avoid redundant data stocks and the related disadvantages (cf. section 7.2). Alternatively, the principle of multiple data management can be used, i.e., the data are stored as copies at different locations.

A disadvantage of this method is the high expenditure for the synchronization of data stocks and the short-term infringement of the integrity by unilateral data changes. Nevertheless, this method is often used in practice because it can easily be implemented and allows for the intermittent autonomous operation of sub-systems.

Figure 9.2 demonstrates the method of a multiple data management at the example of the article item data. These data are required by a WMS as well as a merchandize management system (MMS). In addition to the part which is commonly used by both systems – shaded in grey in the center of Fig. 9.2 – the article master also hold specific data for the WMS and the MMS.

Only the mutually used part builds the interface and has to be synchronized. Generally, the data are synchronized cyclically in times of low activity, mainly at night. Other containers with highly dynamical data have to be synchronized much more frequently, in the extreme case even after each change.

9.2 Classical implementation of a WMS

This section describes some aspects of an implementation by means of a relational database. This method to implement the logical concept of a WMS is typical for currently offered systems. Because of the suggestive structure of relational databases and their plausible query language systems can be implemented within a short time.

9.2.1 Functional structure

A WMS can be implemented in two ways:

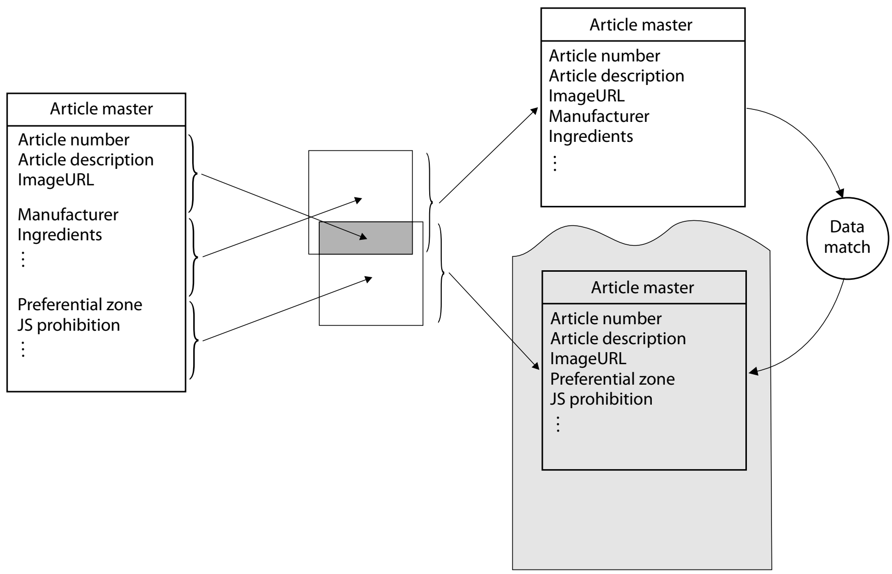


Figure 9.2. Implementation of interfaces by a partially redundant data management

– Mainframe-based architectures (cf. Fig. 9.3), where the WMS runs on a central computer with a database server[4].
– Client-server systems (cf. Fig. 9.4) which bind distributed work stations with specific functions to a database server.

In a mainframe architecture all processes run on the central computer and are operated at terminals with no WMS functions of their own. In client-server systems these processes are distributed completely or partially to the intelligent work stations. The informational tasks are divided as follows:

– Masks can be configured and data represented locally without database access, i.e., by a HTML browser.
– Plausibility checks can be carried out by a client. The correct entry of zip codes or e-mail addresses, for example, can be checked locally. A data transfer from the client to the server and back is an unnecessary communication.
– Integrity checks can be carried out locally when they refer completely to a current operation. For example, when the serial number for an article is entered it can be checked locally that the field for the number of articles

[4] The database operations can also be encapsuled by means of an application server to offer services on a WMS level.

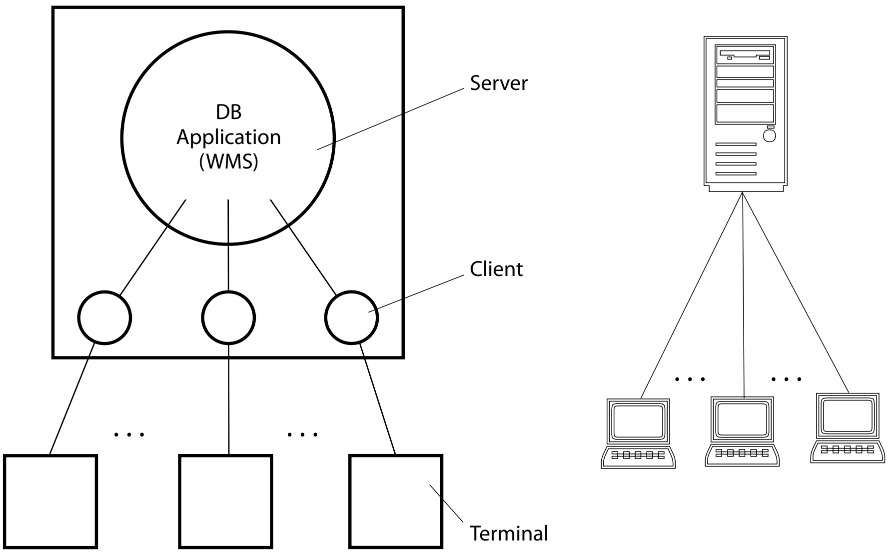


Figure 9.3. Example of a mainframe-based architecture

is not unequal 1. The *uniqueness* of the serial number, however, cannot be checked locally[5].

– Depending on whether data are requested from the database server or not computations can be made locally or just on the server.
– Generally, database operations have to be carried out on the server with the exception of distributed databases which are partially realized by clients (*peer-to-peer solution*).

In practice, it is not always easy to define if a system is based on a mainframe architecture or a client-server solution.

9.2.2 Table structure

The database allows for the structured storage of all data which are necessary for or generated during the operation. This structuring is facilitated by the table concept. Generally, each container shown in Fig. 9.1 can be represented in a database table of its own.

[5] Unless a "serial number" consists of parameters which guarantee for uniqueness like, for example, the millisecond which since 1st January 1970 concatenates with a primary key describing the input terminal.

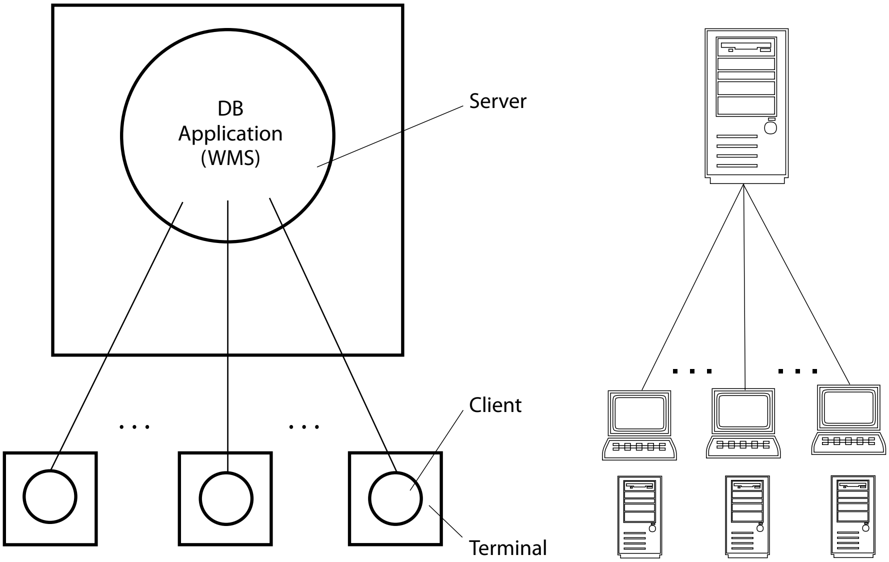


Figure 9.4. Example of a client-server architecture

Before the data can be stored, however, each table has to be structured, i.e., the following storage parameters have to be specified

- what (which)
- where (which table)
- how (which accuracy)
- whereby (which data type)

It makes sense, for example, to store a telephone number as an *integer number*[6] However, the zero as first digit of the area code or special symbols like a plus or a hyphen cannot be stored as an integer data type. For this reason, a telephone number should be stored as a character string where a sufficient number of characters (accuracy) has to be chosen as possible length.

Logically related data like the *ItemData* (cf. Fig. 9.1) are entered line by line into the specified cells of a table. To use relational databases and tables the primary keys for the identification of data lines have to be assigned in a definite way.

A database table *ItemData* could be structured as shown in Table 9.1. In this structure the cell `photo` has a length of 255 characters to enter a URL (see page 233) for the photo. The database type *blob* (Binary Large Object) is one possibility for storing such binary data. The advantage of the URL compared to the blob is that the storage location for the photo data can be freely chosen. Furthermore, multiple references can be made and the database

[6] A data type integer represents a positive or negative integer number.

Table 9.1. Structure of the database table itemData

Name	Type	Length
rId	int	
artno	varChar	20
name	varChar	42
desciption	varChar	255
photo	varChar	255
size	varChar	80
weight	int	
pref.zone	varChar	12
reor.level	int	
act.stock	int	

Table 9.2. Structure of the table stockUnit

Name	Type	Length
rId	int	
serial no	varChar	20
number	int	
itemDataId	int	

does not have to handle large data volumes. A similar method is to describe the data in a separate text file where they can be accessed via a URL. With regard to internationalized descriptions the URL offers many possibilities to manage and represent the data.

In Table 9.1 the field ID `rId` defines a clear rawID as primary key. The article number could not be used as primary key because it would lead to ambiguity at least at the implementation of a client warehouse. Present database systems allow for the automatic generation of primary keys for the single data sets and thus relieve the programmer.

There is a relation $\overset{1}{}\!\!\underline{\quad\quad}\!\!\overset{n}{}$ between ItemData and StockUnit. The ItemData may refer to no, one or several StockUnit tables. This relation is defined by the field `itemDataId` (cf. Table 9.2) that assigns exactly one `rId` of ItemData to *each* line of StockUnit while an `rId` of ItemData can appear in several lines of StockUnit (cf. Fig. 9.5).

The $\overset{n}{}\!\!\underline{\quad\quad}\!\!\overset{m}{}$ association between location and locationCategory can only be built with an auxiliary table, the reference table, also called *map table* (cf. Fig. 9.6).

Each line of the reference table connects a rawID of the table location to a rawID of the table locationCategory and contains a primary key which can be built from the two rawIDs of the respective line.

Without such a reference table the data stock would be highly redundant what would lead to additional memory and runtime requirements.

9.2.3 Securing the logical integrity

Fig. 9.6 shows the structure of the database table locationCategory where the field `characteristic` describes the single zones. The reasonable assignment of a location to a locationCategory has to be ensured by a logistic integrity. Depending on the application each location should or has to be assigned to *exactly one* zone A, B or C. From the logistical point of view a location must not be assigned to several zones but may be allocated to a weight category which is represented by the same field.

Alternatively, the contents of locationCategory can be distributed to several dedicated database tables. This problem is avoided by a distribution with only $\underline{1 \quad n}$ relations. This would impede a statical structure and additional categories *cannot* be generated during operation without having to change the data model.

Another problem regarding the safe logical integrity is described on page 314. When in the database table stockUnit the field `serial number` is occupied the value in the field `number` *has to* be 1.

9.2.4 Generation and query of master data

Data stored in a database are accessed and manipulated by the user in a specific database language (DML[7]). Most database systems support SQL[8] as standardized query language. A user has to know the data model to use SQL for queries and changes of a data stock. The examples describe some typical database operations carried out in SQL.

[7] Data Manipulating Language
[8] Structured Query Language

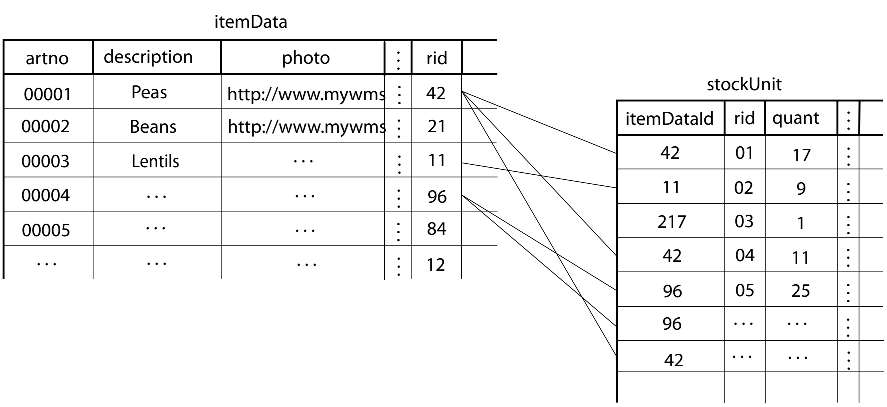


Figure 9.5. Example of a 1 to n relation

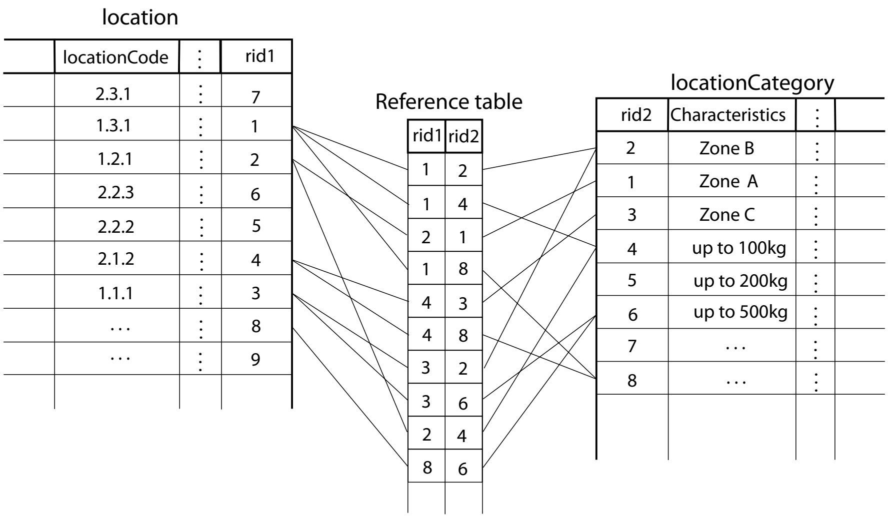


Figure 9.6. Example of a n to m relation

A new data set is entered into the table stockUnit withh the following command where the cell contents are defined by variables. In the example the `stockUnitCount` represents a *sequence* which contains the number of data sets in the table stockUnit:

```
insert into stockUnit
            (rId, serial number, number, itemDataId)
       values
            (stockUnitCount.nextVal,'',11,42);
```

The names of all articles with a weight fewer than ten weight units can be retrieved with

```
select name
       from itemData
       where weight < 10;
```

If all shelves in zone A should be selected this is achieved with the command:

```
select count (*)
       from reference table
       where locationCategory.characteristic = 'Zone A'
       and rid2 = locationCategory.rid2;
```

In the example above the number of selected lines in the reference table corresponds to the number of bin locations which fulfill the desired criterion.

With the function *Stored Procedures* many relational databases offer the possibility to store frequent queries on the server together with the transfer parameters.

9.3 myWMS

The object-oriented programming (OOP) offers many advantages also for the implementation of a WMS. For this reason often object-oriented programming languages[9] are used for more advanced systems. This method has the best effect when not only the programming, but also the analysis and the design of a software are based on it. The focus does not lie on the representation of data structures in a database table, but on an object-oriented analysis. During this phase of the analysis not only the data and their interrelations are described, but also possible *methods* referring to these data. The data are represented in software structures and realized by means of a OOP later on.

Such an OOP solution for WMS was developed in the project myWMS[10]. This section describes the technical concept of myWMS as an example for the consistent use of the OOP technique in WMS. From now on, knowledge about object-orientation and Java is required. But owing to the suggestive structure of this chapter and the chosen examples, the basic principles can be understood without such know-how. Section 9.4 shows the use of this frame for the implementation of a simple WMS.

9.3.1 The basic structure of myWMS

Like an operating system, myWMS provides several basic functions and elementary services for WMS applications. It is a platform and, thus, cannot be operated without specific applications. myWMS creates well-defined communication interfaces to external systems and internal interfaces to the exchangeable modules, the so-called *plug-ins*, which are specified by Java interfaces and allow third parties to integrate own products.

In addition to the pure warehouse management, the myWMS framework also includes a material flow control. myWMS does not deal with aspects of merchandize management, production planning and production control although the framework includes interfaces to such systems.

Since myWMS is not limited to special types of warehouses and logistic processes, its range of application is wide, from manually operated warehouses and automatic warehouses to physically distributed warehouses with heterogeneous structures. Thanks to suitable plug-ins like routing or optimization algorithms the system can easily be adapted to existing warehouses.

[9] According to the general terminology "object-oriented" is abbreviated with OOP.
[10] see http://www.mywms.net

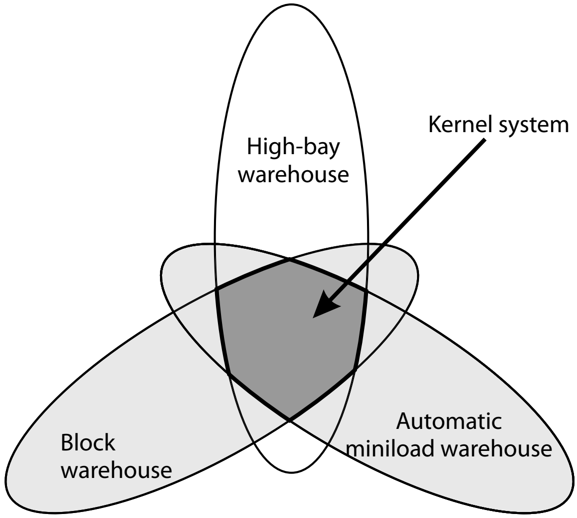


Figure 9.7. As intersection of functions and data structures the core system *simplifies* and *idealizes* different types of warehouses

A plug-in has to be bound to a process – later called *kernel*. A kernel provides basic functions independent of the type of warehouse (cf. Fig. 9.7). myWMS offers a library for frequently used plug-ins and the user can optionally use the original or a modified version or develop his own plug-ins. Thus, a system can have the USP of the user but still benefit from the advantages of a standardized system.

A modular and up-gradable warehouse management system, which is not yet specialized for a certain application, should at least meet the following requirements:

– Independent of platforms: A programming language which allows for the generation of portable codes assures that the system runs on different computer architectures and operating systems.
– Distributability: It should be possible to concurrently run the modules of a WMS on several cooperating computers. The material flow control, for example, could run on another computer than the warehouse management system itself. Since the system is independent of a platform it can be operated on heterogenous computer systems.
– Scalability: Increasing requirements, like a growing throughput rate, for example, or more clients should be met by a WMS by means like distribution, adjustment of modules and parametrization.
– Parametrization: A WMS should allow for the continuous parametrization of its modules. For this purpose the configuration concept should facilitate the persisting of the parameters.

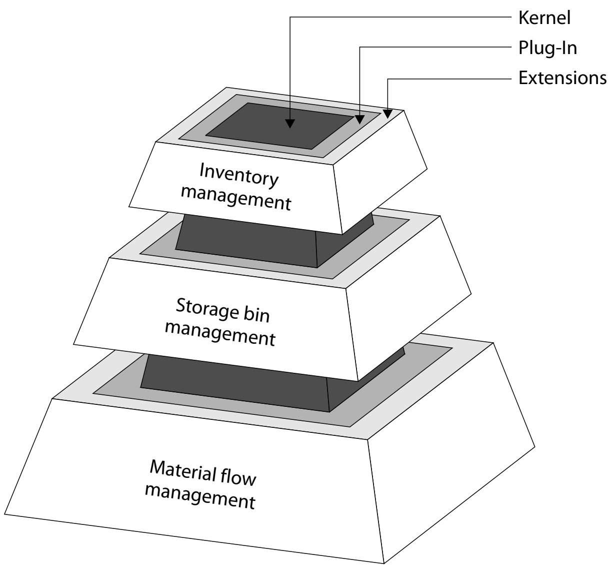


Figure 9.8. The hierarchical model of myWMS

- Releasability: The long-term constancy of the interfaces is an important aspect for the selection of a WMS. The basic software can be exchanged retaining the existing modules and their application-specific extensions as these modules can be exchanged retaining the basic software.
- Web-based operation: The system can be operated with any web browser and is thus independent of platforms. This allows for the integration of the existing IT-infrastructure.
- Independence: No limitation to certain warehouse types, operating means or strategies.
- Upgradability: The system can be adapted to the logistic requirements by generating additional or upgrading existing modules.

myWMS fulfills these requirements by means of the kernel/plug-in technology, a conceptional distribution, the configuration capability and an implementation in Java. myWMS is based on a three-level hierarchy consisting of the inventory management, the location management and the material flow control (cf. Fig. 9.8).

The inventory manager is responsible for the management of item data and article-related operations independent of the location as well as of available and existing quantities of stock units (cf. Fig. 9.1).

The *location manager* manages the current occupancy of the *locations* by unit loads as well as the attributes of the locations.

The *equipment manager* calculates the transport routes, orders the operating means and supervises the operative execution of the transport orders.

Not all systems have to use all three levels. For example, an inventory and location manager can be operated while the bottom level is replaced by a separate material flow computer. In a non-automated warehouse the bottom level, i.e., the equipment manager, may print transport instructions or control radio terminals.

Since the bottom level is operated separately an independent material flow computer without a warehouse function can be built on a myWMS basis.

Each level can also be operated in several instances. A multi-client warehouse, i.e., the provision of several logical warehouse in one warehouse, can be built, for example, by instantiating the two top levels. By means of a multiple instantiation of the bottom level a multiple warehouse, i.e., a warehouse with several physical locations, can be operated. This also allows for the building of multi-client multiple warehouses.

9.3.2 Business objects

The single levels are responsible for managing the respective data stored in the *business objects*. Each business object exactly corresponds to a container in the data model shown in Fig. 9.1. myWMS includes the following business objects:

- **Supply**
- **ItemData**
- **Order**
- **StockUnit**
- **StockUnitCategory**
- **UnitLoad**
- **LoadHandler**
- **Location**
- **LocationCategory**
- **OrderChain**

Thanks to the hierarchical structure single objects can be modelled in detail, e.g., the physical location is defined as **StorageLocation** and the logical one as **Location**. Generally, each logical location has a physical counterpart. The space on the load take-up of a machine is a physical location without a logical equivalent (cf. page 310).

The business objects have been generated consequently according to the OOP principle, i.e., analysis, implementation and test. No attribute can directly be accessed: All attributes are encapsulated and only accessible by so-called accessor methods. Each business object communicates in the above-mentioned way via interfaces. Other methods are encapsulated and not available.

Business objects can be specialized by inheritage. If new attributes are added which were unknown or not considered during the implementation they are realized by key-value codes. This method is generally known as the design pattern *property* and implemented in Java among others as *hashtable*. It is also possible to combine both methods, inheritage and property technique.

The myWMS business object *ItemData* should be adapted to the current application just by derivation because in the basic category only the attributes *itemNumber* and *name* (plain text description) exist, as in every warehouse[11]. Real warehouses may need other attributes like the location of a photo or a minimum inventory (cf. section 2.2). These would then be declared and defined in an *ItemData* upgraded class together with their accessor methods. They are unknown to the kernel and are used by other objects or plug-ins.

For the business object *LocationCategory* can be reasonable to use properties because this allows for the dynamical extension of location groups.

The *comObject*, a category implemented by myWMS, uses both methods. The basic category provides a protected hash table which is filled with every derivation phase. Each derivation provides the corresponding accessor method.

Adequate business processes have to be modelled and implemented to ensure the functioning of a WMS. myWMS can be notified about these business processes by the *PluginInterface*. Another important component is the *SupplyStrategyInterface* as an extension of the *PluginInterface* and serves the *Locations*. The coupling of business processes and business objects is event-driven and based on the *Listener concept*. The business processes do not exchange any status information and and are considered as stateless by myWMS. Typical business processes are

- Storage
- Order-picking
- Retrieval

When a business object is changed all business processes registered as listeners are notified. A business process which receives such a message checks if additional frame conditions are met and, if necessary, performs the next process step. This may require a change of business objects.

myWMS contains some pre-modelled business processes but generally they have to be adjusted to the actual logistic requirements.

9.3.3 Kernel concept

The three levels – inventory management, location management and equipment management – consist of at least one IT process, a so-called kernel. A WMS consists of the kernels of the three levels, the respective plug-ins and the application extensions. The basic principles of myWMS are:

[11] *ItemData* can also use plug-ins via interfaces, such as *ItemDataLockInterface* to manage article blocks.

- Logistic orientation: In contrast to universal databases or application servers myWMS is tailored to the requirements of logistics, especially the warehouse management.
- Flexibility: Since it is limited to core functions and the use of specific plug-ins the system is highly flexible[12].
- Distributions of kernels: The kernels of the different levels can run on different computers. Furthermore, the single levels may consist of several distributed kernels. Section 9.3.1 describes this principle at the example of a multi-client and a multiple warehouse.
- Event-control: To achieve short response times, the event control is stringently used according to the design pattern *listener*. Here, an object registers by callback with one or several other objects.
- Exception handling: Exceptions are unpredictable events during the runtime like the accidental triggering of a sensor signal by the operating staff. Usually, the result of a radio request is returned to the requester in the form of a feedback which describes the success or failure of the action. When the implementation is correct this feedback is handled adequately. The exception handling enforces a transparent further handling what directly improves the quality of the software.
- Integrating active components: The plug-ins are called by the kernel via their methods. They can concurrently carry out actions as independent threads and thus increase the performance.
- Realization of associations: The business objects do not manage any associations but are realized in the kernel. This is one of the prerequisites for the distribution capability.
- Distribution of business objects: If a kernel temporarily distributes a business process to a concurrent process this object is copied with the same object number. Business objects are capable for serialization and are distributed via different computers.
- Assurance of integrity: Although each business object makes elementary integrity checks additional checks can be made with plug-ins. StockUnit, for example, checks if the serial numbers are correctly completed with allowed characters and blanks. With a plug-in it can be checked if these serial numbers already exist.
- Persisting of data: The kernel requests persisting methods which have to be provided by means of a plug-in interface. Some persisting plug-ins are part of myWMS so that data can be secured via the database and the file system can be used by means of serialization. For teaching and training the data can be stored via the same interface without using a database.

According to the above-mentioned concepts the myWMS kernels and the respective plug-ins should best be implemented in Java because it offers:

[12] This limitation does not imply that the kernels are trivial. The open source code reveals the complexity and efficiency of this concept.

– Event and exception handling as well as multithread
– Interfaces to describe abstract classes
– Basic serialization concepts
– Remote Message Invocation (RMI),
– Extensive network capability
– Extensive class libraries for all kinds of problems

Furthermore, Java is the most commonly taught language used by most programmers and software engineers and widely accepted.

9.3.4 Runtime environment

In addition to the kernel and the plug-in library myWMS has a runtime environment, the "standard environment", which offers many useful functions and can also be used for other projects. This environment is divided into the following sections:

- Logging: The logging records different messages which are transferred to exactly one logserver according to their *logLevel* and category. myWMS provides such a logserver as singleton. Thanks to a categorization into
  - executing computer
  - current user
  - date and time
  - object
  - method
  - thread

  the messages can efficiently be analyzed offline.
  - Statistics: Rough statistical data are collected on a separate server, the statistics server, analogous to the recording of log messages.
  - Configuration: In addition to the parametrization the configuration determines which plug-ins are used. When a user logs in at a configuration provider the registered objects are informed about changed configurations. Thus, the configuration can be changed also during the runtime.
  - Clearing: Exceptions which cannot independently be handled by the system are transferred to a *clearing singleton*[13] with a given number of responses or instructions. If a clearing receives such an exception the correct response or acknowledgement of an instruction has to be chosen from outside.
  - Communication: Thanks to the *channel concept* of myWMS communication objects can be transmitted via different physical and logical lines. myWMS implements the transport via the layer of the ISO/OSI protocol stack (cf. section 7.1.1) like the communication via RS232. The protocols

[13] A class which corresponds to the design pattern of the singleton ensures that it is initiated only once.

in the form of plug-ins allow for the serialization and deserialization of the communication objects.
- Event multicasting: Different listener objects can register at a *panel object* which is a listener and thus build a listener group. Events which arrive at the panel object are sent as a copy to all registered listeners. The event multicasting implemented in myWMS can manage up to two listener groups.

The included runtime environment does not necessarily have to be used; only the interfaces have to be considered. Thus, the system can be integrated into existing systems like an own logging.

9.4 Example of a distribution system using myWMS

This section describes a virtual warehouse with typical warehouse management structures, its topology, operating means and logistic processes.

9.4.1 Description of the example

The functions are described at the example of a distribution system for small household appliances. The system consists of a goods receipt and goods issue and the warehouse with an order-picking zone. (cf. Fig. 9.9).

Arriving units are unloaded by stackers and buffered in reserved pick-up zones. After they have been checked and entered into the system, they are brought by stackers to the I-point of the warehouse system (input spur) or into a separate blocking zone for an extensive quality check. On the one hand, complete units are taken from the storage zone and brought with stackers from the output spur to the buffer zones in the goods issue. On the other hand, picking Us are built in front of the high-bay rack and transported analogously to the goods output after completion.

Description of the warehouse system The warehouse (cf. Fig. 9.10) is a simple, automatic high-bay warehouse. On the one hand, it represents typical components and warehouse processes to meet the requirements of the reference system; on the other hand, it is simplified at many points to make it less complex and to outline the most important aspects.

Topology The presented warehouse has 4000 bins and uses Euro pallets (800 mm×1200 mm) as loading aids where the articles are transported and stored homogeneously. It consists of the function areas high-bay rack, pre-storage zone and order-picking terminals.

The high-bay warehouse consists of four aisles with two high-bay racks each. Each aisle is served by a rack feeder. The pre-storage area connects the different function areas. Traversing cars are used for cross-transports.

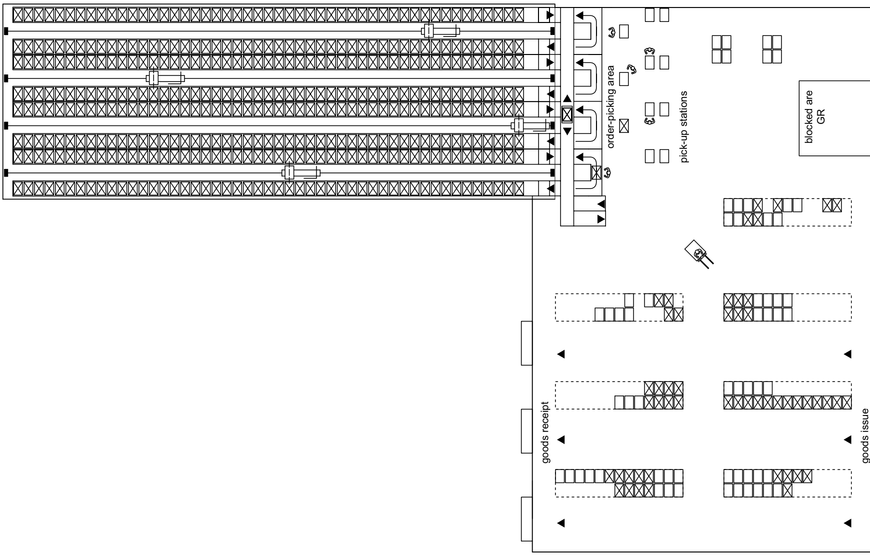


Figure 9.9. The distribution system

Furthermore, the pre-storage area includes the input spur with an I-point, the output spur and two transfer terminals per aisle where the units are transferred between stacker and traversing car. Each of the four order-picking stations consists of the "order-picking U", the space for pickers and pallets to be picked on collecting units. The order-picking U consists of the buffer to be picked from as well as an upstream and downstream transfer terminal as a connection to the traversing cars.

Warehouse technology This form of high-bay warehouse is widely used in practice and represents a classical warehouse technology (cf. Chapter 4). All racks have the same dimensions and are designed for the single-depth lengthwise take-up of standard Euro-pallets. Each of the eight racks consists of 50 sheds in horizontal and 10 sheds in vertical direction, i.e., 500 sheds per rack and a total of 4000 racks. This warehouse does not consider specialities like double-deep storage or racks of different dimensions.

Conveyor technology Each aisle is served by an automatic, rail-bound overhead rack feeder (cf. section 4.2.2). The vehicle cannot change from one aisle into another. All drives are electrical, the energy and control signals are transmitted over contact lines. The load is taken up by a telescopic fork. The most important parameters of the rack feeder are shown in Table 9.3.

The traversing car in the pre-storage area is also rail-bound. However, the energy and control signals are transmitted over trailing cables and the load take-up is a chain conveyor. Since it is a central and decisive element

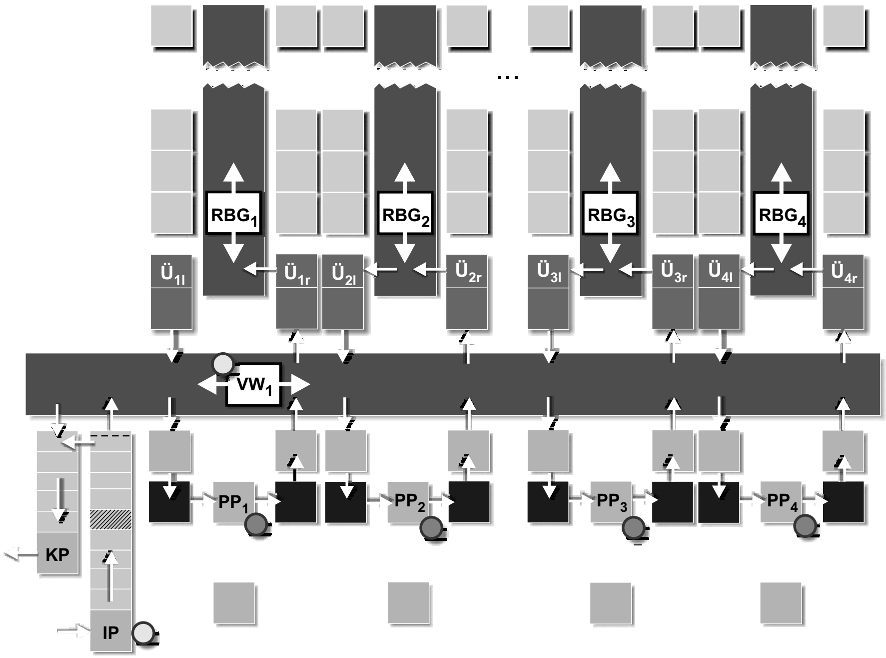


Figure 9.10. The sample warehouse

the traversing car has a high acceleration and speed rate to meet the high requirements.

Two chain conveyors set up in series act as transfer station between pre-storage zone and high-bay rack and offer a buffering capacity of two bins. Each aisle is equipped with two such pairs of chain conveyors: One pair for the input, one for the output. Together with two edge converters the chain conveyor modules build the "order-picking Us". The input and output belts are two driven accumulating roller conveyors. A converter is integrated at the end of the input line which transfers pallets (e.g., in case of a negative contour check) directly to the output line or to the traversing car. Conventional front-end fork lift trucks put pallets on the input line, retrieve pallets from the output line or pick complete pallets. They receive their orders from a separate stacker control system which is no fixed part of this system.

Control technology First, the information for the control of an automatic warehouse is given by the sensors and identification devices. One of these identification media is a barcode label on the loading aid (cf. Chapter 6). This label is applied during the goods receipt process, which takes place outside the system boundary of the reference warehouse and, thus, is not described here. Inside the warehouse each pallet can be identified by two scanning stations. One of these is a stationary scanner at the beginning of

Table 9.3. Technical data of the rack feeder

Parameters	Value	Description
V_x	2,5 m/s	Driving speed (horizontal)
V_y	1,5 m/s	Lifting speed (vertical)
a_{empty}	0,6 m/s^2	Acceleration of the empty rack feeder
a_{loaded}	0,4 m/s^2	Acceleration of the loaded rack feeder
$V_{Fork(empty)}$	1,5 m/s	Telescoping speed of the empty load takeup
$V_{Fork(loaded)}$	1,1 m/s	Telescoping speed of the loaded load takeup

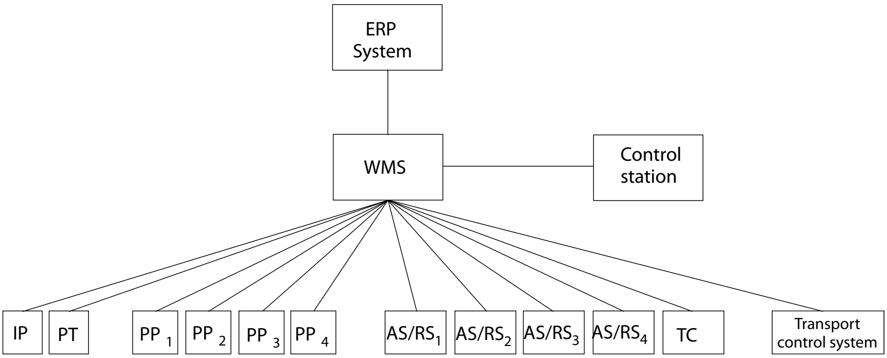


Figure 9.11. System hierarchy (logical view)

the input line and identifies the pallets which enter the system. The second station is a scanner at the traversing car. Furthermore, those pallets which enter the system via the input line are submitted to a contour check.

Various light barriers (e.g., at the end of each accumulating roller conveyor and each transfer station) allow for the tracking and tracing of goods within the warehouse. The order-picking stations are equipped with switches where the picker can acknowledge an order and with terminals which display the article and quantity to be retrieved.

The sensors send all their information via a field bus to superordinate programmable logic controllers (PLC). These PLCs communicate with the superordinate systems via Ethernet and TCP/IP. Depending on the control philosophy this is the warehouse management system or a material flow computer and possible further sub-systems. Each PLC compacts the received sensor information and sends a corresponding telegram to the superordinate system. Each PLC receives control telegrams from the superordinate system which it converts into control commands for subordinate activators. The control architecture is hierarchical and designed as a client/server structure. Figures 9.11 and 9.12 give an overview over the logical and physical architecture of the system.

The conveyor system is controlled by a total of six PLCs: one for each rack feeder, one for the traversing car and one for the remaining pre-storage

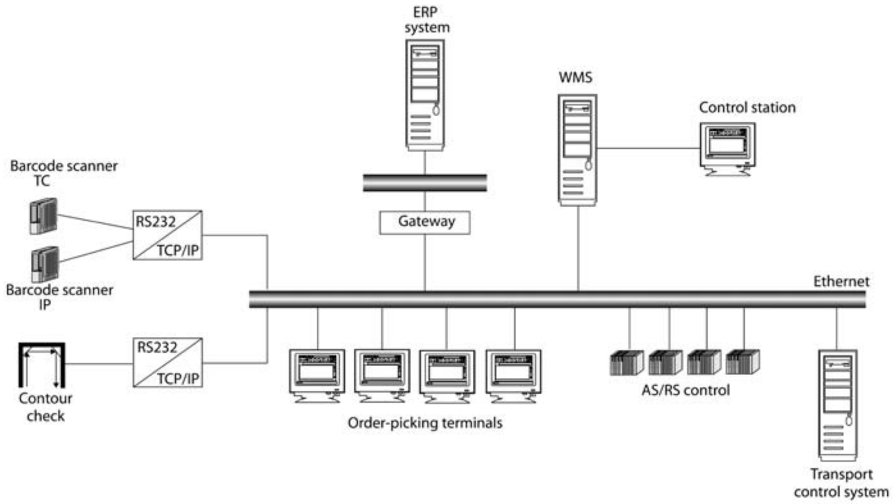


Figure 9.12. System hierarchy (physical view)

zone including the order-picking stations. The PLC layer is the joint interface between subordinate control and superordinate systems.

Material flow The possible routes of a pallet through the reference warehouse are determined by the described topology and the used conveyors. The traversing car plays a central role because it is the sole connection between the function areas and a node[14]. Because of the unidirectional transport at the single transfer stations, the traversing car can either take up a pallet at a station (transfer onto the traversing car) or hand over a pallet at this station (transfer from the traversing car). No transfer station can transfer and receive pallets from the traversing car. Typical routes of a pallet through the warehouse are determined by the logistic functions. Basically these are storage, retrieval and order-picking. Typical routes using a traversing car are:

- **Storage in aisle 4:** identification point → TC → U_{1r} → RF_4 → Storage bin
- **Retrieval from aisle 1:** Storage bin → RF_1 → U_{1l} → TC → Output line
- **Order-picking:** Storage bin → RF_2 → U_{2l} → TC → $Order - pickingU_2$ → TC U_{2r} → RF → Storage bin

Typical routes without a traversing car are:

- **Restorage in aisle 1:** Storage bin → RF_1 → Storage bin
- **Negative contour check:** Input line → Output line

[14] Although transports are bundled on a central device and despite of possible bottlenecks well-dimensioned distribution vehicles are an inexpensive, space-saving and efficient solution.

Logistic processes Conveyors, warehouse technology and controls together with the articles to be managed build the basis for the warehouse management system. The tasks of a WMS are described in Chapter 2. Basic processes closely related to the material flow are:

A) Storage
B) Retrieval
C) Order-picking

The specific characteristics of these processes are roughly described below and determine the requirements on the WMS to be implemented.

Storage When a pallet is transferred at the identification point the unit load is identified by the stationary scanner. The corresponding PLC then sends a telegram with the pallet ID to the WMS where the message triggers the process. First, it is checked if the pallet ID is known. If not, this is an error since the pallet ID should have been sent to the WMS during the goods receipt process together with the article ID. In this case, an order is generated to transport the pallet onto the output line so that it can be discharged and brought to the clearing station. The same happens when the scanner can identify no pallet ID (NO READ).

When the WMS knows the pallet ID, the pallet is moved to the accumulating roller conveyor and to the contour check. If the scanned profile does not correspond to the specifications the pallet is directly discharged from the system via the output line. Otherwise, the goods are booked into the inventory, a storage order is generated and a corresponding storage bin is blocked. This bin should be chosen so that the articles are

A) distributed evenly over all aisles
B) arranged within the aisles according to an ABC zoning.

The even distribution of articles over all aisles improves the accessability, while the ABC zoning grants a quick access on fast-selling articles. If no suitable bins are available an article can principally be stored in each empty bin. The storage process is terminated when the respective load carrier could be booked on the correct storage bin.

Retrieval and order-picking Customer orders are entered into the WMS via an ERP system and stored in a queue. A customer order generally consists of several items. If a position cannot be fulfilled because there are not enough units of the desired articles in store the complete order is rejected and a message is sent to the ERP. Otherwise, the WMS includes the order into the queue which is handled according to the FIFO principle. Based on the customer orders, the WMS generates retrieval orders and selects the unit loads to be retrieved. It has to make sure that the unit loads are chosen according to the following – possibly conflicting – goals:

A) The number of pallets to be picked should be as small as possible, i.e., unit loads should be retrieved directly. Broken pallets should be considered as well.
B) The unit load with the oldest storage date should be retrieved at first.
C) The rack feeders should be loaded as evenly as possible.

Pallets which cannot be picked are removed from the warehouse by traversing cars and output lines. When the traversing car picks up the pallet at the transfer station it makes a scan to ensure that the pallet is the right one. If this is not the case the pallet is brought to the clearing station. Otherwise, when the pallet arrives at the output line the stacker control system is requested to bring this pallet to the buffer in the goods issue determined in the order.

If a position cannot be completed with complete pallets articles have to be picked. In this case, the pallet identified as suitable by the WMS is brought by the traversing car from the high-bay warehouse to the specified picking station. There, the picker is informed on his terminal how many units he has to retrieve and to place onto the provided pallet. By pressing a switch he informs the WMS that the picking order is complete. The WMS then defines a new storage bin for the pallet. It furthermore checks if all positions of the current order have been completed. If this is the case the WMS instructs the stacker control system to remove the order pallet to the respective buffer in the goods issue where picked and directly retrieved unit loads are consolidated and packed for shipment.

9.4.2 Topology structure

The topology of a warehouse which consists of storage bins, machines and their connections is set up in Java using the kernel methods. The storage bins are represented by objects of the class *StorageLocation* where different access methods can be chosen (e.g., LIFO, FIFO, RandomAccess) depending on the requirements. The location of the bins is of no importance for the WMS and thus not shown.

Objects of the class *HandlingScope* represent a working area for the machines which connect the objects of the class *StorageLocation*. Each *LoadHandler* which represents a machine operates at a time on exactly one *HandlingScope*. The following code shows the principle topology structure:

```
// Structure of an aisle side

EquipmentKernelInterface ek = ... // Connection to kernel

int x=50;                  // Aisle depth : 50 bins
int y=10;                  // Aisle height : 10 bins
int i,j;                   // Run variables
long tid;                  // Transaction handle

StorageLocation sl;        // temporary SL for ALL bins
```

```
HandlingScope hs;           // temporary HS for total aisles

String HandlingScopeName =...// Name of the aisle

...
  tid = ek.beginTransaction(); // Start of a transaction
  hs = ek.getHandlingScope(HandlingScopeName, tid);
                             // existing handlingScope
                             // incl. requesting Loadhandler
  for ( i = 0 ; i < x ; i++ ) {
    for ( j = 0 ; j < y ; j++ ) {
      try{
        sl = ek.createStorageLocation(tid);
        sl.setName(getBinName(i,j));  // getBinName provides
                                       // individual name
        sl.setStorageType(StorageLocation.LIFO);
        sl.setDepth(1);              // one UnitLoad per bin
        ek.setStorageLocation(sl,tid); // sl to kernel
        hs.addStorageLocation(sl);  // Connection hs with sl
      }
      catch (...) {
        ...
      }
    }
  }

  try {
    ek.setHandlingScope(hs,tid);  // hs return to kernel
  }
  catch (IntegrityException iex){
    ...
  }
  try {
    ek.commit(tid);              // Carry out transaction
  }
  catch (TransactionException tex){
    ...
  }
...
```

This code fragment requires a *HandlingScope* as well as a *LoadHandler* including its load take-up consisting of a *StorageLocation*. In this example the business objects are instantiated by the 500 *StorageLocations* according to the `create` methods of the kernel, here `ek.createStorageLocation(tid)`. Each business object can be addressed by its name which is assigned by the method `sl.setName(name)` of the *StorageLocation*.

Then the type “LIFO” is assigned to the storage bin with the method `sl.setStorageType(StorageLocation.LIFO)` as well as the capacity with the method `sl.setDepth(1)`. The type LIFO was chosen for racks considering a future extension for a multiple-depth storage. With such an extension it would be suitable to set the shelf depth with the method `sl.setDepth(n)`.

Furthermore, the access strategy would have to be adjusted to the racks since it cannot be guaranteed that the desired loading aid can directly be accessed. The kernel is informed about this adjusted access strategy by the plug-in method.

For further management, the method `ek.setStorageLocation(sl,tid)` is used to transmit the newly generated and configured *StorageLocation* to the kernel. Finally, the rack is connected to the HandlingScope by `hs.addStorageLocation(sl)`. The method `hs.addStorageLocation(sl)` ensures a bi-directional material flow (stacker $\Leftrightarrow$ bin).

With this method all other bins, in addition to the 3500 sheds, are created and allocated to at least one HandlingScope. The transfer stations between rack feeder and traversing car, for example, are connected to two HandlingScopes. Since the material flow at the transfer stations is bi-directional the methods `hs.addPutAway(sl)` or `hs.addRetrieval(sl)` have to be used for the *HandlingScopes*.

The line `ek.beginTransaction()` starts a transaction which is then managed by its transaction handle `tid`. This transaction is stopped only when `ek.commit(tid)` is called. Thus, the generated objects are persisted and become available for other kernel methods.

9.4.3 Plug-In – Routing

The route from a source, e.g., a shelf, to a sink, e.g., an order-picking station, is calculated by a plug-in which fulfills the `RouteStrategyInterface`. With the class `RouteStrategyFirstMatch`, myWMS provides a simple routing plug-in which always finds a route by means of excursion.

In practice it will do to adjust this class to the requirements according to the inheritance principle. Often, it is sufficient to overwrite the method `findWay()` of the class `RouteStrategyFirstMatch` with an algorithm which meets the warehouse requirements. `findWay()` is called by the method `route()` of the plug-in interface.

Alternatively, it is possible to fulfill the interface directly with an own class without deduction. In addition to the inherited interfaces[15] just the method `route()` has to be implemented.

```
public interface RouteStrategyInterface
                 extends PluginInterface  {

  OrderChain[] route(final OrderChain[] orders, long tid)
    throws TransactionException;

}
```

Pairs of sources and sinks are transferred to the method `route()` and `route()` computes a route for each of these pairs. Source and sink are a

[15] PluginInterface and its predecessor ConfigurationClientInterface.

StorageLocation and are transmitted to an otherwise empty OrderChain. The route is also returned to the caller as OrderChain, but with exact details consisting of the single StorageLocations connected by HandlingScopes.

`route()` is the multiple application of the mathematical function r which operates on the transport network with the node l and the edges in the HandlingScope h:

$$r(l_1, l_2) = \begin{cases} null & l_1 = l_2 \\ null & \text{there is no route} \\ l_1, (h, l)*, h, l_2 & \text{otherwise} \end{cases}$$

The function reveals that there is not always a route between two StorageLocations.

The represented plug-in has to be registered at the kernel. For this purpose, a corresponding entry is made into the configuration file. The configuration manager instantiates the plug-ins and integrates them into the kernel. Figure 9.13 shows the interrelation of the different components:

– The kernel provides a plug-in interface with pre-defined method names, their arguments and feedback parameters.
– A plug-in has to fulfill this interface.
– The plug-in can be extended by inheritance.
– The configuration determines which plug-in is used by the kernel.

This method has to be observed by all plug-in interfaces. myWMS offers an extensive library of predefined plug-ins by which a simple WMS can be set up.

9.4.4 Communication

The communication between myWMS and the subordinate actuators and sensors, respectively their combination into sub-systems, is a basic part of the equipment kernel which controls the material flow. During this communication the kernel sends and receives objects while the connected devices receive and send byte streams of different formats.

The basic class of all communication objects is the class ComObject which provides basic methods such as the property management. The design pattern “property” with the two methods `setProperty()`and `getProperty()` facilitates the implementation of different classes derived from ComObjects.

The class DirectedMessage is directly derived from ComObject and with the protected methods `setProperty()` and `getProperty()`provides public functions to set and get the following attributes:

– `sender` generates the instance.
– `receiver` receives the message.
– `functionCode` describes the type of message.

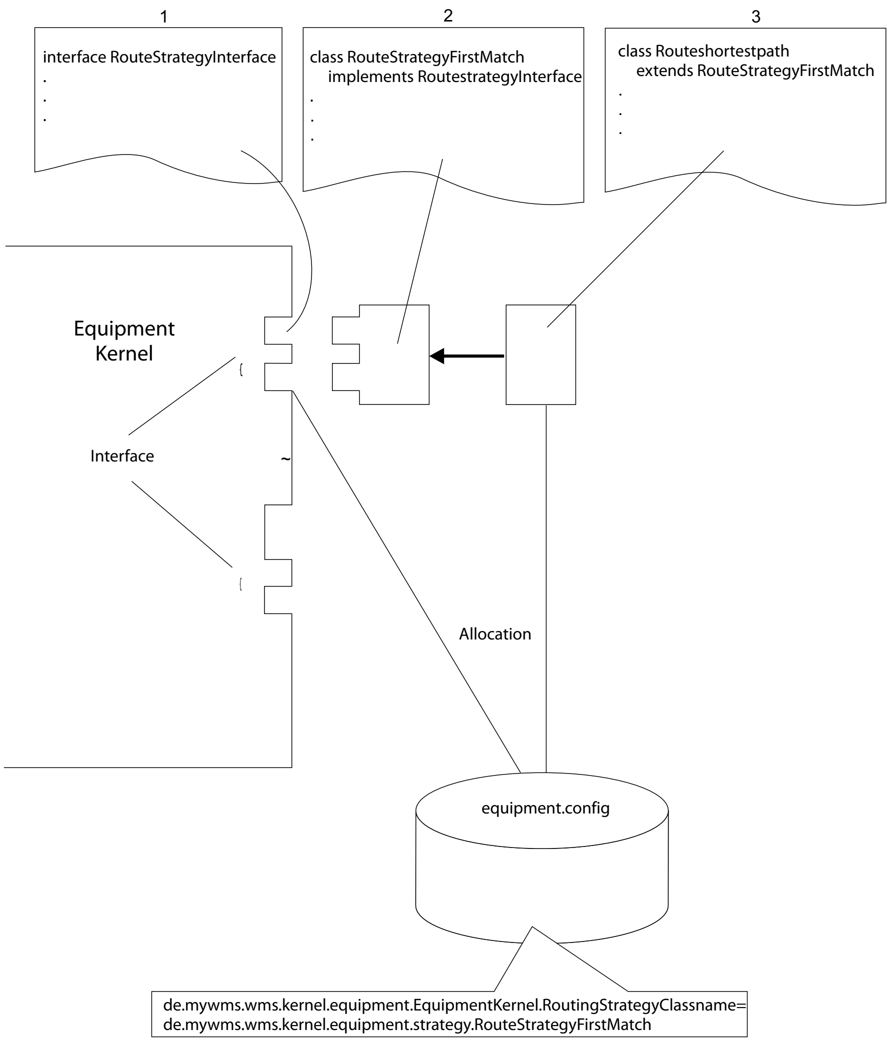


Figure 9.13. Integration of a plug-in into the kernel by configuration

– `sequenceNumber` is a counter for the generation of primary keys.

If a `functionCode` is given which is known to myWMS a special object is generated. Figure Fig. 9.14 shows possible structures in a class diagram. It is clearly defined how specialized communication objects have to be handled in myWMS. Communication objects with a known `functionCode` cannot be evaluated by myWMS but are designed for application-specific extensions.

The class MfcTransportOrder extends DirectedMessage and generates orders for transport facilities. The required new attributes are not designed as

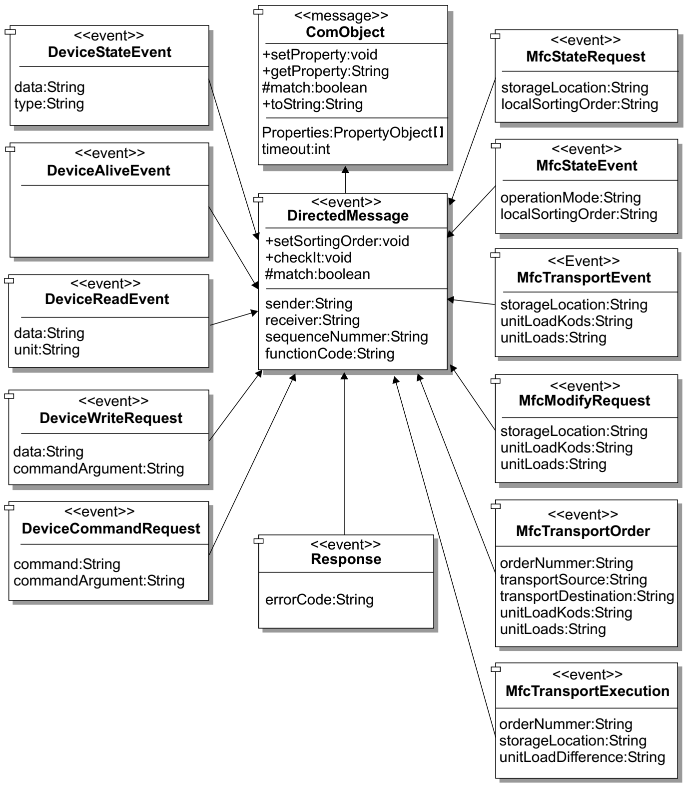


Figure 9.14. Diagram of the communication classes

elements of this derivation level but stored in the property structure of the root class. These new attributes are:

– `transportSource` the starting point of the transport.
– `transportDestination` the final point of the transport.
– `orderNumber` a consecutive number for the transport.
– `unitLoads` the number of transport units.

The single steps of a transport order are represented in MfcTransportOrder in the form of a routing strategy and communicated to the executing machine by the protocol channel method.

The communication objects are exchanged with the environment via the channel (cf. section 9.3.4). The channel is responsible for the data transport and uses the protocol object to serialize the communication objects or to construct them from the byte streams.

```
public interface Protocol {
```

```
  ...

  public byte[] eval(ComObject comObject, ...)
                throws ComException;

  public ComObject[] eval(byte[] bytes, ...);

}
```

Each protocol has to comply with the above mentioned protocol interface. The most important method is `eval` which has to be implemented twice by overloading.

– During a serialization the first parameter transferred to eval is a ComObject. According to the protocol to be implemented the ComObject is converted into an array of bytes (byte[]) and returned to the caller.
– During the deserialization the first parameter transferred to eval is a byte array. The byte array is the section of a received stream. For the construction of a ComObject it does not matter whether it is incomplete, contains data for one or several ComObjects, incomplete parts of the next or is faulty. Complete byte sequences are converted into ComObjects and returned in an array. Incomplete byte sequences have to be buffered until the next eval request.

myWMS includes several protocols with different targets. There are classes for the communication with a choice of AutoID devices while other classes represent protocols for the data exchange with material flow computers. The class XMLProtocol is a universally suitable protocol which uses the flexibility of the XML meta language and all XML tools.

```
<message type="WMSmessage">
  <header sequence="0005" receiver="PLC3"
          sender="WMS7" />
  <MfcTransportOrder unitLoads="0001"
          orderNumber="3"
          transportSource="1.2"
          transportDestination="1.7" />
</message>
```

This example shows a serialized MfcTransportOrder object which was serialized by an instance of the XMLprotocol and sent in this form to a transport facility (LoadHandler). The identifiers for the transport source and sink correspond to the “point of view” of the transport device.

This section describes basic techniques at the example of simple distribution systems. It was pointed out that a WMS includes more than just one *data model*. Warehouse management systems are complex systems which have to meet different requirements and thus are a challenge for software engineers. In recent years this science has developed into a valuable engineering discipline - especially thanks to the use of object-oriented techniques and suitable *usable* methods and tools.

The future goal is to fulfill different requirements in a high quality and without individual reprogramming by means of suitable methods and standards. In addition to WMS which are based on relational databases the presented concept is *one* possibility to reach this goal.

Abbreviations

AGTS	Automatic guided transport system
AGV	Automatic guided vehicle
ANSI	American National Standards Institute
AS/RS	Automatic storage and retrieval system
Auto-ID	Automatic identification
CEP	Courier, express and parcel(Services)
DC	Double cycle of the RF = loaded trip and return
DDL	Data definition language
DLL	Dynamic link library
DNS	Domain Name Service
EDP	Electronic data processing
ERP	Enterprise resource planning
EDI	Electronic data interchange = data exchange format
ePC	Electronic product code
EP	Euro pallet
ETSI	European Telecommunication Standards Institute
FIFO	"First-in-first-out" = Retrieval of articles according to the storage date (oldest article first)
FM	Frequency modulation
FPS	Flexible production system
GC	General contractor
GI	Goods issue
GLN	Global location number, corresponds to ILN
GR	Goods receipt
HBW	High-bay warehouse
HF	High frequency
HTML	Hypertext markup language = standardized language for the World Wide Web
I-point	Identification point
IP	Internet protocol, IP address: Number to identify a computer in the Internet
ID	Barcode identification number
ILN	International location number

ISDN	Integrated services digital network, method to transmit voice and data messages via a telephone line
ISO	International Organisation for Standardization
JIT	Just in time
KPI	Key performance indicator
LAN	Local area network, internal network
LHM	Ladehilfsmittel, z.B. Karton, Behälter oder Palette
LIFO	"Last-in-first-out" = order of retrieval of an article of the same kind according to the storage date (last article first)
LF	Low frequency
LTU	Load take-up
MDT	Mobile data terminal (without online connection to the WMS)
MFC	Material flow controller
MIS	Merchandize information system
NSU	Number of shipping unit
OCR	Optical-character recognition
OSI	Open systems interconnection
PIN	Personal identification number
PL	Procedural language
PLC	Programmable logic controller
QA	Quality assurance
ONS	Object name service
QU	Quantity unit = e.g., piece, m^3, m^2, m, kg
PDA	Production data acquisition
PLC	Programmable logic controller to control the movements of the RF and the conveyor system. The PLC is connected online with the MFC.
PPC	Production planning and control
PU	Pick-up unit = single-type crate or box gripped from a UL
RDT	Remote data transmission
RF	Rack feeder
RFID	Radio frequency identification
RFC	Remote function call = Call of a function from a remote PC or host
PML	Physical markup language
SC	Single trip of the RF = one loaded and one unloaded trip
SCS	Stacker control system
SKU	Stock keeping unit
SQL	Structured query language
SSCC	Serial Shipping Container Code
SU	Sales unit
TC	Traversing car
TCP/IP	Transmission control program / Internet protocol
TSU	Transport unit

TSP	Travelling salesman problem
TTS	Transport, transshipment and storage
UHF	Ultra high frequency
UL	Unit load as storage and transport unit
UPS	Uninterrupted power supply
URL	Uniform resource locator = Internet address
WAN	Wide area network or goods acceptance
WMS	Warehouse management system
WCS	Warehouse control system
μC	Micro-controller

Bibliography

[1] *ISO/IEC 9798-2 Informationstechnik - IT-Sicherheitsverfahren - Authentifikation von Instanzen, Teil 2: Mechanismen auf Basis von Verschlüsselungsalgorithmen.* Berlin: Beuth-Verlag, July 1999

[2] Abraham Silberschatz, Peter B. G.: *Operating System Concepts.* 3rd edn. Addison-Wesley, 1991

[3] Alicke, Knut: *Modellierung und Optimierung mehrstufiger Umschlagsysteme.* Karlsruhe, Universität Karlsruhe, Dissertation, October 1999

[4] Arnold, Dieter: *Materialflusslehre.* 2nd edn. Verlag Vieweg & Sohn Braunschweig Wiesbaden, 1998

[5] Balzert, Helmut: *Lehrbuch der Softwaretechnik — Software-Entwicklung.* Spektrum Akademischer Verlag, Heidelberg, Berlin, 1996

[6] Balzert, Helmut: *Lehrbuch der Software-Technik – (Lehrbücher der Informatik); Bd. 1. Software-Entwicklung.* 2nd edn. Spektrum, Akad. Verl., Heidelberg Berlin, 2000

[7] Bauer, Friedrich L.; Goos, Gerhard: *Informatik : Eine einführende Übersicht - Erster Teil.* 3rd edn. Berlin, Heidelberg, New York: Springer, 1982

[8] Bittner, Fred; Bulga, Martin: Barcodeleser mit Fuzzy-Decodern. In: Jünemann, Reinhardt (Hrsg.); Wölker, Martin (Hrsg.): *Automatische Identifikation in Praxis und Forschung.* Verlag Praxiswissen, 1996

[9] Bretzke, Wolf-Rüdiger: Die Eignung der Transaktionskostentheorie zur Begründung von Make-or-Buy-Entscheidungen. In: Pfohl, Hans-Christian (Hrsg.): *Logistikforschung – Entwicklungszüge und Gestaltungsansätze.* Erich Schmidt Verlag, Berlin, 1999, S. 335–363

[10] Buchmann, Johannes: *Einführung in die Kryptographie.* 2. Berlin, Heidelberg: Springer, 2001

[11] CCG, Centrale für Coorganisation: *Cross Docking zwischen Handel und Industrie – Eine ECR-Anwendungsempfehlung der ECR-Initiativen Deutschland und Österreich.* March 2000

[12] CCG, Centrale für Coorganisation GmbH: *EAN 128.* 3rd edn. 2001

[13] Copco, Atlas: *Druckluft Antriebspraxis.* 1996

[14] *eLogistics-Facts 1.0.* Market Research Service Center und Deutsche Post eBusiness GmbH eCommerce Services in Zusammenarbeit mit Fraunhofer Institut Materialfluss und Logistik, November 2001

[15] Dietze, Günther; Figgener, Olaf: Studie Lagerverwaltungssysteme und ihr Leistungsprofil. In: *F+H Fördern und Heben* 51 (2001), Nr. 1-2, S. 59–60

[16] Dietze, Günther; Figgener, Olaf; Spee, Detlef: Studie Lagerverwaltungssysteme und ihr Leistungsprofil. In: *F+H Fördern und Heben* 50 (2000), Nr. 12, S. 18–19

[17] *DIN 19226-1 – Leittechnik - Regelungstechnik und Steuerungstechnik - Allgemeine Grundbegriffe.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, February 1994

[18] *DIN 19233 – Leittechnik - Prozeßautomatisierung - Automatisierung mit Prozeßrechensystemen, Begriffe.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, July 1998

[19] *DIN 44300 Teil 2: Informationsverarbeitung; Allgemeine Begriffe.* Deutsches Institut für Normung e.V., Beuth-Verlag, 1988

[20] *DIN 55405 – Begriffe für das Verpackungswesen.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, 1988-1993

[21] *DIN 55429 – Schachteln aus Karton, Vollpappe oder Wellpappe, Bauarten Ausführungen Lieferformen.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, June 1985

[22] *DIN 60000 – Informationsverarbeitung - Mathematische Zeichen und Symbole der Schaltalgebra.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, November 1985

[23] *DIN 8402 – Qualitätsmanagement und Qualitätssicherung - Begriffe.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, 1991

[24] Ernst, Hartmut: *Grundlagen und Konzepte der Informatik.* Vieweg Verlag, Braunschweig, Munich, 2000

[25] Fliege, N.: *Digitale Mobilfunksysteme.* Stuttgart: B.G. Teubner Verlag, 1996

[26] Franzke, Ulrich: Formale und endliche Melodiesprachen und das Problem der Musikdatencodierung. In: Boroda, M. (Hrsg.): *Musikometrica 5.* Brockmeier, Bochum., 1993

[27] Goyal, Amit: Savant Guide / Auto-ID Center, Massachusetts Institute of Technology. Cambridge, MA, April 2003 (15). – Forschungsbericht

[28] Göpfert, Ingrid: Trends bei Logistik-Kennzahlen. In: *Fördern und Heben* 46 (1996), Nr. 4, S. 254–258

[29] Grosseschallau, W.: *Materialflussrechnung.* Springer, Berlin Heidelberg, 1984

[30] Gudehus, Timm: *Grundlagen der Kommissioniertechnik.* Verlag W. Girardet, Essen, 1973

[31] Gudehus, Timm: *Logistik: Grundlagen, Strategien, Anwendungen.* Springer, Berlin Heidelberg, 1999

[32] Hafner, Sigrid (Hrsg.): *Industrielle Anwendung evolutionärer Algorithmen.* R. Oldenbourg Verlag, Munich, 1998

[33] Hansen, Hans-Günter; Lenk, Bernhard: *Codiertechnik.* Ident Verlag Veuss, 1996

[34] Hartmann, Rainer: *Untersuchung absolut codierter, magnetischer Maßstäbe zur Positionierung spurgeführter Fördermittel.* Dortmund, University of Dortmund, Dissertation, 1998

[35] Hein, Mathias: *TCP/IP – Internet-Protokolle im praktischen Einsatz.* 4th edn. International Thomson Publishing, Bonn, 1998

[36] Heistermann, J.: *Genetische Algorithmen.* B.G. Teubner Verlag, Stuttgart, Leipzig, 1994

[37] *DIN EN 61131-1 – Speicherprogrammierbare Steuerungen - Teil 1: Allgemeine Informationen.* Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin, August 1994

[38] Jünemann, Reinhardt: *Materialfluß und Logistik – Systemtechnische Grundlagen mit Praxisbeispielen.* Springer, Berlin Heidelberg, 1989

[39] Jünemann, Reinhardt; Beyer, Andreas: *Steuerung von Materialfluß und Logistiksystemen.* 2nd edn. Springer, Berlin Heidelberg New York, 1998

[40] Jünemann, Reinhardt; Schmidt, Thorsten: *Materialflußsysteme – Systemtechnische Grundlagen.* 2nd edn. Springer, Berlin Heidelberg New York, 1999

[41] Jungnickel, Dieter: *Optimierungsmethoden - Eine Einführung.* Springer, Berlin, Heidelberg, New York, 1999

[42] Kauffels, Franz-Joachim: *Moderne Datenkommunikation — Eine strukturierte Einführung.* International Thomson Publishing, Bonn, 1996

[43] Kinnebrock, W.: *Optimierung mit genetischen und selektiven Algorithmen.* R. Oldenbourg Verlag, Munich, Vienna, 1994

[44] Krampe, Horst (Hrsg.); Lucke, Hans-Joachim (Hrsg.): *Grundlagen der Logistik – Einführung in Theorie und Praxis logistischer Systeme.* 2nd edn. Huss-Verlag, Munich, 2001

[45] Kuhn, A.; Hellingrath, B.: *Supply Chain Management – Optimierte Zusammenarbeit in der Wertschöpfungskette.* Springer, Berlin Heidelbarg, 2002

[46] Kuhn, A.; Pielok, Th.; Sümpelmann, A.: Prozesskettenanalyse - Instrumente für die strategische Planung. In: *Jahrbuch der Logistik 1993* (1993), S. 160 ff.

[47] Lange, Volker; Brachetti, Christoph: *Mehrweg-Transport-Systeme.* Verlag Praxiswissen, Dortmund, 1997

[48] Lenk, Bernhard: Identifikation mittels Strichcode. In: Jünemann, Reinhardt (Hrsg.); Wölker, Martin (Hrsg.): *Automatische Identifikation in Praxis und Forschung.* Verlag Praxiswissen, Dortmund, 1996

[49] Lenk, Bernhard: *Handbuch der automatischen Identifikation.* Monika Lenk Fachbuchverlag, 2000

[50] Lenk, Bernhard: *2D-Codes, Matrixcodes, Stapelcodes, Composite Codes, Dotcodes.* Monika Lenk Fachbuchverlag, 2002

[51] Linti, Frank: Beitrag zum fachgerechten Einsatz von Barcodeetiketten in Förder- und Lagertechnik. In: Jünemann, Reinhardt (Hrsg.); Wölker, Martin (Hrsg.): *Automatische Identifikation in Praxis und Forschung.* Verlag Praxiswissen, Dortmund, 1996

[52] Lämmerhirt, E.-H.: *Elektrische Maschinen und Antriebe.* Carl Hanser Verlag, Munich, Vienna, 1989, 1989

[53] Lolling, Andreas: Fehlerraten verschiedener Kommissionersysteme in der Praxis. In: *Zeitschrift für Arbeitswissenschaft* 56 (2002), Nr. 1-2, S. 112–115

[54] Martin, Heinrich: *Praxiswissen Materialflussplanung.* Verlag Vieweg & Sohn, Braunschweig, Wiesbaden, 1999

[55] Mäusl, R.: *Digitale Modulationsverfahren.* Heidelberg: Hüthig Verlag, 1985

[56] Meinke, H; Gundlach, F.W.: *Taschenbuch der Hochfrequenztechnik.* 5. Berlin, Heidelberg: Springer, 1992

[57] Merz, Hermann: *Elektrische Maschinen und Antriebe.* 1. VDE Verlag, Berlin, Offenbach, 2001

[58] Messmer, Hans-Peter: *PC-Hardwarebuch.* 4th edn. Bonn, Paris: Addison-Wesley, 1997

[59] Mettler Toledo: *Logistiksysteme für Gewichts- und Volumenerfassung*

[60] N.N.: Gerolsteiner positioniert sich neu – Automatische Chargenverfolgung. In: *Logistik Heute* 23 (2001), June, Nr. 6, S. 20–22

[61] Papageorgiou, Markos: *Optimierung: statische, dynamische, stochastische Verfahren.* 2nd edn. R. Oldenbourg Verlag, Munich, 1996

[62] Pfohl, Hans-Christian: *Logistiksysteme – Betriebwirtschaftliche Grundlagen.* 6th edn. Springer-Verlag, Berlin Heidelberg, 2000

[63] Plessmann, K.W.: *Handbuch zum VDI-Bildungswerk GmbH Seminar: Lasten- und Pflichtenheft als Vorbereitung und Voraussetzung der Entwicklung von Software.* Stuttgart, October 2000

[64] Radtke, Axel: *Beitrag zur Entwicklung optimierter Betriebsstrategien für Sortiersysteme*, University of Dortmund, Dissertation, 2000

[65] Raiffa, Howard: *Einführung in die Entscheidungstheorie.* R. Oldenbourg Verlag, Munich, Vienna, 1973
[66] Reichmann, Thomas: Kennzahlen. In: Bloech, Jürgen (Hrsg.); Ihde, Gösta B. (Hrsg.): *Vahlens großes Logistiklexikon*, Verlag Franz Vahlen, Munich, 1997, S. 425–426
[67] Reissenweber, Bernd: *Feldbussysteme.* Oldenbourg Verlag, Munich, Vienna, 1998
[68] Riessler, Frank: Etikettendruck: Warum Thermo und Thermotransfer? In: Jünemann, R. (Hrsg.); Wölker, M. (Hrsg.); Krämer, K. (Hrsg.): *Basis der Unternehmenprozesse - Identifikationstechnik.* Umschau Zeitschriftenverlag, Frankfurt, 1999
[69] Rohling, H.: *Einführung in die Informations- und Codierungstheorie.* Teubner Studienbücher: Elektrotechnik, B.G. Teubner Verlag, Stuttgart 1995
[70] Rosenbaum, Oliver: *Das Barcode–Lexikon.* edition advanced, bhv Verlags GmbH, 1997
[71] Sauer, Jürgen: *Multi-Site Scheduling – Hierarchisch koordinierte Ablaufplanung auf mehreren Ebenen.* Oldenburg, University of Oldenburg, Habilitation, January 2002
[72] Scheffler, M.: *Grundlagen der Fördertechnik – Elemente und Triebwerke.* Verlag Vieweg & Sohn, Wiesbaden, 1994
[73] Scheffler, M.; Feyrer, K.; Matthias, K.: *Fördermaschinen – Hebezeuge, Flurförderzeuge, Aufzüge.* Verlag Vieweg & Sohn, Wiesbaden, 1997
[74] Spee, Detlef: LVS-Auswahl per Internet - Kosten sparende Entscheidungshilfe. In: *Hebezeuge und Fördermittel* 41 (2001), June, S. 282–283
[75] Stroustrup, Bjarne: *Die C++ Programmiersprache.* Addison-Wesley-Longmann, Bonn, 1998
[76] Tanenbaum, Andrew: *Moderne Betriebssysteme.* 2nd edn. Carl Hanser Verlag, Munich, Vienna, 1995
[77] Tanenbaum, Andrew S.: *Computernetzwerke.* Prentice-Hall Verlag, Munich, 1997
[78] ten hompel, Michael: *Lagerverwaltungssysteme – Das Objekt-Ebenenmodell der Lagerverwaltung.* Praxiswissen, Dortmund, 1999
[79] Tompkins, James A. (Hrsg.); Smith, Jerry D. (Hrsg.): *The Warehouse Management Handbook.* Tompkins, Raleigh, 1998
[80] Tompkins, James A.; White, John A.; Bozer, Yavuz A.; Frazelle, Edward H.; Tanchovo, J.M.A.; Trevino, Jaime: *Facilities Planning.* 2nd edn. Wiley, New York, 1996
[81] United States Patent Office. *Pat.-No. 2612994: Classifying Apparatus and Method.* October 1949
[82] *VDI 2319 – Übersichtsblätter Stetigförderer – Angetriebene Rollenbahn.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, July 1971
[83] *VDI 2326 – Übersichtsblätter Stetigförderer – Gurtförderer für Stückgut.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, June 1979
[84] *VDI 2340 – Ein- und Ausschleusungen von Stückgütern – Übersicht, Aufbau und Arbeitsweise.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, March 1997
[85] *VDI 2395 – Gleislose Fahrzeugkrane.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, September 1989
[86] *VDI 3312 – Sortieren im logistischen Prozeß – Entwurf.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, December 1998
[87] *VDI 3586 – Flurförderzeuge – Begriffe, Kurzzeichen.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, March 1996

[88] *VDI 3590 Blatt 1 – Kommissioniersysteme – Grundlagen.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, April 1994
[89] *VDI 3590 Blatt 1 – Kommissioniersysteme – Systemfindung.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, July 2002
[90] *VDI 3591 –Transportleitsysteme.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, December 1998
[91] *VDI 3619 – Sortiersysteme für Stückgut.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, March 1983
[92] *VDI 3968 Blatt 1-6 – Sicherung von Ladeeinheiten.* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, 1994
[93] *VDI 4422 – Elektropalettenbahn (EPB) und Elektrotragbahn (ETB).* Verein Deutscher Ingenieure, Beuth-Verlag, Berlin, September 2000
[94] *VDMA-Einheitsblatt 15276: Datenschnittstellen in Materialflusssystemen.* Verband Deutscher Maschinen- und Anlagenbau, Beuth-Verlag, Berlin, 1994
[95] VOGEL, Johannes: *Elektrische Antriebstechnik.* 6th edn. Hüthig GmbH Heidelberg, 1998
[96] W. DIFFIE, M. H.: *New Directions in Cryptography.* IEEE Transactions on Information Theory, November 1976
[97] WALK, ELDOR: Aktuelle Situation der RFID Standards für die Logistik. In: *ident Jahrbuch 2004.* Rödermark: ident Verlag, 2004, S. 48–53
[98] WEBER, Jürgen: *Logistikkostenrechnung.* 2nd edn. Springer, Berlin Heidelberg, 2002
[99] WEBER, Michael: *Verteilte Systeme.* Spektrum Akademischer Verlag, Heidelberg, Berlin, 1998
[100] WEBER, Wolfgang: Data Matrix ECC200. In: *ident - Das Forum für Automatische Datenerfassung, Jahrbuch 2002/2003.* Umschau Zeitschriftenverlag, Frankfurt, 2001
[101] WECK, Gerhard: *Prinzipien und Realisierung von Betriebssystemen.* Teubner Verlag, Stuttgart, 1982
[102] WESSELMANN, Friedhelm: *Untersuchungen zur Weiterentwicklung automatischer Blocklagersysteme mit gekuppelten Rollpaletten*, University of Dortmund, Dissertation, July 1997
[103] WIESNER, Werner: *Der Strichcode und seine Anwendungen.* Verlag Moderne Industrie, Landsberg/Lech, 1990
[104] WILL, D.: *Einführung in die Hydraulik und Pneumatik.* Verlag Technik, Berlin, 1983
[105] WIRTH, Niklaus: *Compilerbau.* 4th edn. B. G. Teubner, Stuttgart, 1984
[106] ZEHNDER, C. A.: *Informationssysteme und Datenbanken.* 5th edn. B. G. Teubner Stuttgart, 1989
[107] ZIEGLER, H.-J.: *Computergestützte Transport- und Tourenplanung.* Expert Verlag, Ehningen bei Böblingen, 1988
[108] ZIMMERMANN, H.-J.: *Operations Research - Methoden und Modelle.* Vieweg Verlag, Braunschweig, Wiesbaden, 1987

Index

System requirements

THE MYWMS DEMO CD IS A BOOT CD!
When the CD is inserted and the computer is switched on, a Linux operating system on the CD is booted.

hardware: System requirements for the myWMS demo systems are a x86 PC, min. 512 MB memory and 1 GHz CPU. A higher memory and computing capacity of the CPU will increase the system performance considerably.

Boot version: This CD is bootable. After the CD drive has been activated as boot device (to be set in the BIOS) the system is booted directly from the CD without access to the hard disk. When the CD is removed the PC runs as usual.

Linux: The CD holds a JRE (Jave Runtime Environment), version 1.4.1, which don't need to be installed.

Windows: Under Windows a JRE (Jave Runtime Environment) version 1.4.n has to be installed on the PC. It can be obtained at java.sun.com as packed or executable file.